BOOK OF ABSTRACTS

International Conference on Open Data: Open Data Challenges and Opportunities in Times of Crisis and Growth (ICOD 2022)
November 28th – December 2nd 2022, Zagreb, Croatia

ICOD 2022

International Conference on Open Data:
Open Data Challenges and Opportunities in Times of Crisis and Growth
November 28th – December 2nd, 2022, Zagreb, Croatia

Organized by:

Faculty of Law, University of Zagreb

and

Faculty of Agriculture, University of Zagreb
Faculty of Geodesy, University of Zagreb
Faculty of Electrical Engineering and Computing, University of Zagreb
Faculty of Organisation and Informatics, University of Zagreb
Delft University of Technology, Netherlands
University of Aegean, Greece

within the consortium of the project:
Twinning open data operational (TODO) – 857592

Title: International Conference on Open Data (ICOD 2022): Book of abstracts

Editors:
Filip Varga (fvarga@agr.hr), Faculty of Agriculture, University of Zagreb, Croatia; Centre of Excellence for Biodiversity and Molecular Plant Breeding (CroP-BioDiv), Zagreb, Croatia
Petra Đurman (pdurman@pravo.hr), Faculty of Law, University of Zagreb

Publisher: Faculty of Law, University of Zagreb

ISBN: 978-953-270-167-8

Web: icod2022.pravo.hr

Citation: Varga, F., Đurman, P. (Eds.) (2023). International Conference on Open Data (ICOD 2022): Book of abstracts. Faculty of Law, University of Zagreb, Zagreb, Croatia, pp. 182.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 857592-TODO

ICOD 2022

International Conference on Open Data:
Open Data Challenges and Opportunities in Times of Crisis and Growth

November 28th – December 2nd 2022, Zagreb, Croatia

International Conference on Open Data: *Open Data Challenges and Opportunities in Times of Crisis and Growth* was organized within the Horizon2020 TODO (*Twinning Open Data Operational*) project at the University of Zagreb, Croatia from November 28 – December 2, 2022. International conference was organized by the Faculty of Law, University of Zagreb as task leader (T5.5) within the project work package 5 *Dissemination and outreach*, in cooperation with other project partners.

Under the main title *Open Data Challenges and Opportunities in Times of Crisis and Growth*, this interdisciplinary scientific conference covered a wide range of topics on legal, political, social, financial, and technical challenges in the provision and use of open data, in general and in specific sectors (for example, traffic data, geospatial data, etc.). The conference took place from Monday, November 28, to Friday, December 2, 2022, physically at the University of Zagreb (Trg Republike Hrvatske 14) and the Faculty of Law (Trg Republike Hrvatske 3), with online streaming of all sessions and panels via Zoom platform. **Therefore, the conference was successfully organized in a hybrid mode (offline and online), in English as the language of the conference.**

The conference encompassed 8 sessions, 5 panel discussions, two keynote lectures by eminent professors and two workshops. On 28th November, a pre-conference day, a panel on fiscal transparency and a workshop on high-value datasets were held. The official opening of the conference followed on 29th November at the Dubrovnik Hotel. Introductory and welcoming speeches were given by the Vice-Rectors of the University of Zagreb, Associate Professor Anamarija Musa and Professor Jurica Pavičić, Professor Almin Đapo, Dean of the Faculty of Geodesy, University of Zagreb, Professor Elizabeta Ivičević Karas, Vice-Dean of the Faculty of Law, University of Zagreb, Assistant Professor Ana Kuveždić Divjak, project coordinator from the Faculty of Geodesy, University of Zagreb, and Hrvoje Meštrić, the Head of the Directorate for Science and Technology of the Ministry of Science and Education. The opening of the conference was followed by a panel on open data policies in a comparative perspective and a keynote speech by Professor Marijn Janssen from Delft University in the Netherlands. Two more sessions followed on the possibilities of open data for improving the transparency of public administration and for the development of the private sector. On 30th November, two sessions were held - on open data ecosystems and applications based on open data, as well as two panels – one on legal aspects of open data in general, and another on open data in the context of GDPR. The fourth day of the conference (1st December) started with a keynote lecture by Professor Joep Crompvoets from the Catholic University of Leuven (Belgium), followed by a session and two panels on the topic of geospatial data. On the last day of the conference, 2nd December, a workshop was held on mapping and visualization of open data, followed by a session on media and cultural aspects of open data, as well as session containing poster presentations. The final speech with the conclusions and summaries of the conference was delivered by TODO team leader of the Faculty of Law and

president of the Scientific Committee of the conference, Professor Anamarija Musa, Vice-rector of the University of Zagreb.

In addition to scientists from different disciplines, the conference hosted professionals from public authorities, civil and private sector organizations, as well as open data users and enthusiasts, students and other interested stakeholders. The conference eventually gathered more than 300 participants, with 34 papers presented. Out of 34 papers, 23 were related to the TODO consortium, with at least one author coming from TODO partner institution. The main presenters of the papers were coming from 7 EU countries (Croatia, The Netherlands, Greece, Austria, Germany, Italy, Poland) and 4 non-EU countries (India, Indonesia, China, USA).

Editors,

Petra Đurman, University of Zagreb Faculty of law

Filip Varga, University of Zagreb Faculty of Agriculture

ICOD 2022

International Conference on Open Data:
Open Data Challenges and Opportunities in Times of Crisis and Growth
November 28th – December 2nd 2022, Zagreb, Croatia

Organizing Committee

Petra Đurman, PhD
Faculty of Law, University of Zagreb, Croatia
Martina Erdelić, PhD
Faculty of Transport and Traffic Sciences, University of Zagreb, Croatia
Marko Jurić, PhD
Faculty of Law, University of Zagreb, Croatia
Bia Mandžuka
Faculty of Transport and Traffic Sciences, University of Zagreb, Croatia
Andrea Miletić
Faculty of Geodesy, University of Zagreb, Croatia
Tereza Rogić Lugarić, PhD
Faculty of Law, University of Zagreb, Croatia
Filip Varga, PhD
Faculty of Agriculture, University of Zagreb
Centre of Excellence for Biodiversity and Molecular Plant Breeding (CroP-BioDiv), Croatia

Scientific Committee

Charalampos Alexopoulos, PhD
Department of Information and Communication Systems Engineering, University of the Aegean, Greece
Željko Bačić, PhD
Faculty of Geodesy, University of Zagreb, Croatia
Igor Čavrak, PhD
Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia
Tihomir Katulić, PhD
Faculty of Law, University of Zagreb, Croatia
Ana Kuveždić Divjak, PhD
Faculty of Geodesy, University of Zagreb, Croatia
Bastiaan van Loenen, PhD
Faculty of Architecture and The Built Environment, Delft University of Technology, Netherlands
Euripidis Loukis, PhD
Department of Information and Communication Systems Engineering, University of the Aegean, Greece
Anamarija Musa, PhD
Faculty of Law, University of Zagreb, Croatia
Dragica Šalamon, PhD
Faculty of Agriculture, University of Zagreb
Miroslav Vujić, PhD
Faculty of Transport and Traffic Sciences, University of Zagreb, Croatia
Frederika Welle Donker, PhD
Faculty of Architecture and The Built Environment, Delft University of Technology, Netherlands
Anneke Zuiderwijk, PhD
Faculty of Technology, Policy and Management at Delft University of Technology, Netherlands
Nikolina Žajdela Hrustek, PhD
Faculty of Organization and Informatics, University of Zagreb, Croatia

Monday 28 November 2022

Location: Faculty of Law, University of Zagreb

11:00-12:00	Panel	**Open Budgets and Tax Transparency: Towards Fiscal Openness**	Chair: **Tereza Rogič Lugarić** (UNIZG LAW)
13:30-17:00	Workshop	**Identification of High-Value Dataset determinants: is there a silver bullet for efficient sustainability oriented data-driven development?**	Chair: **Charalampos Alexopoulos** (AEGEAN)

Tuesday 29 November 2022

Location: Hotel Dubrovnik, Ljudevita Gaja 1, Zagreb

9:00-10:00		**Registration**	
10:00-11:00	**Opening session**	**OPEN DATA CHALLENGES AND OPPORTUNITIES IN TIMES OF CRISIS AND GROWTH**	Chair: **Anamarija Musa** (UNIZG LAW)
	Opening speeches	**Jurica Pavičić**, Vice-rector for International and Inter-institutional Cooperation of the University of Zagreb **Almin Đapo**, Dean of the Faculty of Geodesy, University of Zagreb **Elizabeta Ivičević Karas**, Vice-dean for science, Faculty of Law, University of Zagreb **Ana Kuveždić Divjak**, Project Coordinator, Faculty of Geodesy, University of Zagreb **Hrvoje Meštrić**, director of the Directorate for Science and Technology, Ministry of Science and Education, Croatia	
11:00-12:00	**Panel**	**ADVANCEMENTS IN OPEN DATA POLICIES: A COMPARATIVE PERSPECTIVE**	Chair: **Anamarija Musa** (UNIZG LAW)
12:00-12:30	**Keynote lecture**	Marijn Janssen **„AI for Enhancing Open Data“**	

13:30-15:00	**Session I**	**OPEN GOVERNMENT DATA: NEW OPPORTUNITIES FOR GOVERNMENT AND BUSINESS** Shaharudin, van Loenen, Janssen **„Identifying business models of open data intermediaries: A review“** Loutsaris, Maratsi, Lachana, Ali, Alexopoulos, Charalabidis **„Open Government Data and NFTs. A proposed solution for Governments“** Rizun, Ciesielska, Alexopoulos, Saxena, Papageorgiou **„Open Government Data (OGD) in European educational programs curriculum: current state and prospects“** Welle Donker, Kuveždić Divjak **„Open Data Developments, Opportunities and Challenges in Europe: lessons learned from Open Data best practices in 2022“**	Chair: **Dragica Šalamon** (UNIZG AGRI)
15:30-17:00	**Session II**	**OPEN DATA ENHANCING TRANSPARENCY** Alexopoulos, Rizun, Saxena **„How do OGD initiatives affect transparency? A post-analysis on developed indices“** Kapoor, Jehling **„Analyzing the impacts of Open Data Portals on Urban Development and Innovations in German Cities“** Varga, Kević, Hrustek **„Transparency in agricultural land lease by local government“** Rogić Lugarić, Musa, Đurman, Klemenčić **„Open budgets: open data as a tool for fiscal and tax transparency“** Kević, Miletić, Kuveždić Divjak, Welle Donker **„Readiness for Re-use of Open Government Data for Solving Societal Problems Based on Public Participation: Kindergartens in the City of Zagreb“**	Chair: **Frederika Welle Donker** (TUDELFT)

Wednesday 30 November 2022

Location: Faculty of Law, University of Zagreb

9:30-11:00	**Session III**	**OPEN DATA ECOSYSTEMS** Calzati **„Back to Roots: Reconsidering Data as a Resource to Introduce a Modulating Approach to Open Data“** Supinajaroen, van Loenen, Korthals Altes **„NCORS Open Data Ecosystem: beyond open data! “** Ali, Alexopoulos, Charalabidis **„Data interoperability and the open data ecosystem: roles and research areas“** Schwoerer **„Designing Open Government Data Programs for Usability: The Impact of Usability Attributes on Open Data Use“**	Chair: **Anneke Zuiderwijk van Eijk** (TUDELFT)
11:30-13:00	**Session IV**	**OPEN DATA APPLICATIONS** Tonec Vrančić, Vujić, Bukvić **„The Analysis of Available Open Data in the EU for the Purpose of Increasing the Safety of Railway-Level Crossings“** Erdelić, Carić, Erdelić, Mardešić **„Identification of Features for Trajectory Segmentation According to the Transport Mode“** Aracri, Feretti, Motta, Ferreira, Bibuli, de Pascalis, Odetti, Bruzzone, Caccia **„Open Science in Marine Robotics“** Basu **„Digital India Mission: Revolutionising Open Data Mechanisms since 2015“** Mandić, Bubnić, Matić, Bišćan, Garašić, Vujić **„Selection of its solutions for improving the quality of the traffic system in the city of Zagreb based on user requirements“**	Chair: **Igor Čavrak** (UNIZG FER)
14:00-15:30	**Session V**	**LEGAL CHALLENGES TO OPEN DATA** Mudrić **„Open data, data mining and personal data in law enforcement environment“**	Chair: **Anamarija Musa** (UNIZG LAW)

		Katulić, Musa, Lončar Dušanović **„Privacy and personal data challenges for open data“** Matanovac Vučković **„Right in databases and open data policies - convergence or conflict“** Musa, Habazin, Katulić **„Open legislation: comparison of legal portals in three countries“** Jurić, Musa, Rogić Lugarić **„Health data: reconciling open data benefits with the GDPR“**	
16:00-17:00	**Panel**	**OPEN DATA AND PRIVACY PROTECTION: THE GDPR EFFECT REVISITED** **Marko Trošelj**, Personal Data Protection Agency of the Republic of Croatia **Marija Bošković Batarelo**, Parser Compliance, Zagreb **Darja Lončar Dušanović** **Anamarija Musa**, Faculty of Law, University of Zagreb	Chair: **Tihomir Katulić** (UNIZG LAW)

Thursday 1 December 2022

Location: Faculty of Law, University of Zagreb

9:30-10:00	**Keynote lecture**	Joep Crompvoets **„Is Open Data a hype or mainstream? “**	
10:00-11:00	**Panel**	**OPEN DATA ADVANCES AND CHALLENGES IN GEOSPATIAL SOCIETY** **Ivica Skender**, GDi **Tomislav Ciceli**, DGU **Frederika Welle Donker,** TU Delft **Almin Đapo, Faculty of Geodesy, UNIZG**	Chair: **Željko Bačić** (UNIZG GEOD)
11:30-13:00	**Session VI**	**GEOSPATIAL REASONING WITH OPEN DATA** Spinoza Andreo, Welle Donker, Calzati **„Assessing an open spatial data**	Chair: **Vesna Poslončec Petrić** (UNIZG GEOD)

		infrastructure from a user participation perspective: A qualitative exploratory research with OpenStreetMap“ Welle Donker, van Loenen, Poslončec Petrić **„Open Spatial Data Infrastructure Active Learning and Teaching Methods in Practice“** Kuveždić Divjak, Kević **„Engaging with Open Data through Visualisation and Communication: The Role and Possibilities of Cartography“** Varga, Kuveždić Divjak, Šalamon **„Towards development of a tool for the automated assessment of the spatial accuracy of nature observation datasets“**	
14:00-16:30	**Panel**	**INTERNATIONAL COOPERATION IN GEOSPATIAL DOMAIN: EGYPT-CROATIA** **Amr H.A. Ali,** Benha University, Faculty of Civil Engineering, Cairo, Egypt **Hany M.A. Ayad**, Alexandria University, Faculty of Engineering, Alexandria, Egypt	Chair: **Željko Bačić** (UNIZG GEOD)

Friday 2 December 2022

Location: Faculty of Law, University of Zagreb

9:30-10:00	**Workshop**	**OPEN DATA MAPPING: INCREASING THE IMPACT OF OPEN DATA THROUGH VISUALISATION AND COMMUNICATION**	Chair: **Ana Kuveždić Divjak** (UNIZG GEOD)
10:00-11:30	**Session VII**	**OPEN DATA, NEW MEDIA AND CULTURAL HERITAGE** Papageorgiou, Alexopoulos, Loukis **„Mapping the Scientific Research on Data Journalism and Open Data“** Devi Nur Afni, Wijaya **„Twitter Sentiment Analysis on the Implementation of One Data Indonesia with Semi Supervised SVM“** Gelo Čolić **„Media reporting in the context of data verifiability“**	Chair: **Anamarija Musa** (UNIZG LAW)

		Shankar **„History for Non Historians: Analyzing the Effectiveness of Various Natural Language Processing Models Used in Social Media“** Lemić, Aigner **„Openness and Usability of Big Data of the Past and Time Machine Project – Digital Transformation of the European Cultural Heritage“**	
12:30-13:00	**Poster Session**	Klasinc **„Open, Useful and Reusable Data (OURdata) Policy Impact on Startup Innovation“** Zhu **„Accessing Open Data on NFT“**	Chair: **Anamarija Musa** (UNIZG LAW)
13:30-14:30	**Closing session**	**OPEN DATA CHALLENGES AND OPPORTUNITIES IN TIMES OF CRISIS AND GROWTH**	Chair: **Anamarija Musa** (UNIZG LAW)

SESSION I

OPEN GOVERNMENT DATA: NEW OPPORTUNITIES FOR GOVERNMENT AND BUSINESS

IDENTIFYING BUSINESS MODELS OF OPEN DATA INTERMEDIARIES: A REVIEW

Ashraf Shaharudin[1]*, Bastiaan van Loenen[1], Marijn Janssen[2]

[1] Faculty of Architecture & the Built Environment, Delft University of Technology, Netherlands
[2] Faculty of Technology, Policy & Management, Delft University of Technology, Netherlands

*correspondence E-mail: A.A.BinAhmadShaharudin@tudelft.nl

Keywords: open data; intermediaries; infomediaries; business models; revenue; value

1. Introduction

Open data has many potential benefits including stimulating innovation, enhancing accountability and transparency, and improving the reproducibility and dissemination of research (Janssen et al., 2012; Uhlir and Schröder, 2007; Zhu et al., 2019). However, there are various shortcomings in the current open data initiatives such as the mismatch between the supply and demand of open data, the lack of appropriate software to process data, and confusion regarding data licenses (Johnson et al., 2017; van Loenen et al., 2021). The role of open data intermediaries is considered important to address these weaknesses. Open data intermediaries facilitate the use of and access to open data (Chattapadhyay, 2014; González-Zapata and Heeks, 2015) and build connections among open data stakeholders (Mayer-Schönberger and Zappia, 2011). It follows that open data intermediaries are regarded as one of the priority areas in open data research (Davies and Perini, 2016).

Nevertheless, in-depth studies on open data intermediaries are scarce. Within the limited studies, they are found to face several challenges that may undermine their potential contribution to other open data stakeholders. For example, lack of financial planning (Flores, 2020), over-reliance on volunteers (Reggi and Dawes, 2016), and difficulty in securing data experts (Andrason and van Schalkwyk, 2017). Some of these challenges are associated with the lack of development of open data intermediaries' business models (Kitsios et al., 2021; Reggi and Dawes, 2016). Before any research-based development of their business models can be carried out, a clear view of the existing business models is needed. Given this, the objective of this extended abstract is to review existing business models of open data intermediaries from the academic literature through a systematic literature review (SLR).

Section 2 provides a brief background on the concept of a business model. Section 3 describes the research method. Section 4 presents the findings. Last but not least, Section 5 discusses the findings and proposes considerations for future studies.

2. Background: What is a business model?

There are various definitions of a business model. For example, Timmers (1998) defined it as "an architecture of the product, service and information flows, including a description of the various business actors and their roles; a description of the potential benefits for the various business actors; a description of the sources of revenues". Twenty years later, Afuah (2018) defined it as

"the set of activities that [a business] performs to build and use resources to generate, deliver, and monetize benefits (embodied in products and services) to customers".

Osterwalder (2004) identified nine components of a business model, which were then refined and developed into a business model canvas (Osterwalder and Pigneur, 2010). The nine components are (1) key partners, (2) key activities, (3) key resources, (4) value proposition, (5) customer relationships, (6) channels, (7) customer segments, (8) cost structure, and (9) revenue streams. Meanwhile, Al-Debei and Avison (2010) identified four components of a business model: (1) value proposition (products/services offered), (2) value architecture (technological architecture and organisational infrastructure, (3) value network (relationships with businesses and customers), and (4) value finance (cost, pricing methods, and revenue structure). It can be seen that the components identified by Osterwalder and Pigneur, (2010) and Al-Debei and Avison (2010) are similar and the difference is only a matter of specificity.

Despite various interpretations of a business model, most scholars agree that it contains at least three main elements (as summarised by Afuah, 2018; Andreini and Bettinelli, 2017; Voigt et al., 2017), namely, (1) value proposition (potential benefits for the customers), (2) value creation (methods deployed by companies to deliver the value proposition to customers), and (3) value capture (payments, not necessarily in monetary form, from customers to companies). As it is beyond the scope of this extended abstract to define or refine what is a business model, we adopted the three elements (value proposition, value creation, and value capture) as a guide in conducting our review.

3. Research method

We followed the eight steps of the SLR process by (Xiao and Watson, 2019). First, we formulated the problem that we wanted to achieve from the SLR. In our case, we want to answer: what are the business models of open data intermediaries in the literature? Second, we developed the review protocol, as presented in this section. Third, we searched for the literature. Fourth, we screened for inclusion by reviewing the title and abstract. Fifth, we assessed the quality of each piece of literature by reviewing the full text. Sixth, we extracted data from the literature. Seventh, we analyzed the data. Lastly, we reported the findings.

We searched for relevant publications in three academic databases, namely Scopus, Web of Science (WoS), and Google Scholar. We conducted the literature search on September 21, 2022. We used the search terms shown in **Table 1**. Although the scope of this extended abstract is not limited to open government data but open data as a whole, we also included the term "open government data" since the academic sub-area of open government data has gained tremendous interest over the years, resulting in much literature in this area. We included the term "infomediaries" and "infomediary" in our searches since our initial literature scanning shows that it is often used as a synonym for data intermediary. We also included the term "intermediation" to capture literature that uses it instead of "intermediary". Besides the terms "business model" and "business models", we also included the terms "revenue" and "value" since both are keywords closely associated with business models.

Table 1. Search terms (Boolean operator OR across rows and AND across columns)

Boolean operator	**AND**		
OR	open data	intermediaries	business model
	open government data	intermediary	business models
		infomediaries	revenue
		infomediary	value
		intermediation	

In total, there were 35 publications compiled from the three databases (**Table 2**). We removed eight duplicated publications and a publication with no author's information in the first filtering stage, giving us 26 publications. We then removed 19 irrelevant publications (publications that are not about open data intermediaries or business models) based on the title and abstract, and three non-English language publications, leaving us with four publications. Based on the content of each publication, one of them is found irrelevant to the objective of this extended abstract as it is on the business models of open data initiatives of which open data intermediaries are only one of the elements of the business model. In the end, three publications were selected: Janssen and Zuiderwijk (2014), Magalhaes et al. (2014), and Germano et al. (2016).

Table 2. Search strategy and number of results for each database

Database	**Search in**	**Results**	**Notes**
Scopus	title, abstract, keywords	22	N/A
WoS	title, abstract, author keywords, and Keywords Plus	11	N/A
Google Scholar	title	2	Google Scholar only allows terms searched either in the title or in the whole publication. The latter will give about 965,000 publications, hence, the search is only done in the title.

4. Results

Janssen and Zuiderwijk (2014) studied the business models of "infomediaries" in the Netherlands that are driven by open data and social media through 12 cases. They adopted the concept of a business model by Al-Debei and Avison (2010). However, in their analysis, they only focused on the value proposition and identified six business models:

i. Single-purpose apps: Apps that process one type of open data and present it visually
ii. Interactive apps: Single-purpose apps that allow users to add content such as ratings and feedback

iii. Information aggregators: Apps that integrate open data from multiple sources
iv. Comparison models: Apps that aggregate and compare open data from various sources
v. Open data repositories: Portals that publish open data
vi. Service platforms: Platforms that allow the searching, importing, cleansing, processing, and visualisation of open data

In the same year, Magalhaes et al. (2014) studied the business models of commercial reuse of open government data by 500 firms based in the United States. Like Janssen and Zuiderwijk (2014), they adopted the concept of business model by Al-Debei and Avison (2010) but focused only on the value proposition in their analysis. They identified three business models:

i. Enablers: On the supply side, enablers provide public agencies with the services to collect, manage, and publish open data. On the user side, they offer products and services for users to aggregate data from different sources
ii. Facilitators: They simplify and promote access to open data such as by repackaging and republishing data and providing Application Programming Interfaces (APIs)
iii. Integrators: They use open data to complement their internal/private data to augment their business capabilities

A couple of years later, Germano et al. (2016) studied the business models of seven open government data intermediaries in Brazil. They did not clarify which interpretation of business model they adopted. They identified three business models based on the source of revenue:

i. Consultancy services
ii. Sponsorship to brands that want to advertise on their platforms
iii. Products' subscription

5. Discussion

The business models identified from the publications reviewed do not cover all the three main elements considered key in business models, namely value proposition, value creation, and value capture. In particular, business models identified by Janssen and Zuiderwijk (2014) and Magalhaes et al. (2014) only represent the value proposition whereas business models by Germano et al. (2016) only represent the value capture. A holistic view of existing business models of open data intermediaries in terms of what value is offered (value proposition), how the value is delivered (value creation), and how the value is compensated (value capture) is needed to propose research-based development of their business models.

Future studies should consider identifying business models of open data intermediaries that capture all three main elements of business models. Besides, since the three studies reviewed are all country-specific, future studies should consider looking at different geographical scopes to support the generalisability of the current open data intermediaries' business models. Due to the limited studies of open data intermediaries' business models in academic literature, future studies should consider utilising grey literature including use case catalogues (e.g., opendataimpactmap.org) to identify open data intermediaries and subsequently identify their business models.

Acknowledgements: The authors acknowledge the financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955569 – Towards a Sustainable Open Data Ecosystem (ODECO) as well as from TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Afuah, A. (2018). Business Model Innovation: Concepts, Analysis, and Cases. Routledge. Available at: https://search.ebscohost.com/

Al-Debei, M. M., Avison, D. (2010). Developing a unified framework of the business model concept. European Journal of Information Systems 19(3): 359–376. https://doi.org/10.1057/ejis.2010.21

Andrason, A., van Schalkwyk, F. (2017). Opportune Niches in Data Ecosystems: Open Data Intermediaries in the Agriculture Sector in Ghana (SSRN Scholarly Paper No. 2949722). Social Science Research Network. https://doi.org/10.2139/ssrn.2949722

Andreini, D., Bettinelli, C. (2017). Business Model Innovation: From Systematic Literature Review to Future Research Directions. Springer International Publishing. https://doi.org/10.1007/978-3-319-53351-3

Chattapadhyay, S. (2014). Access and use of government data by research and advocacy organisations in India: A survey of (potential) open data ecosystem. Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance, 361–364. https://doi.org/10.1145/2691195.2691262

Davies, T., Perini, F. (2016). Researching the emerging impacts of open data: Revisiting the ODDC conceptual framework. The Journal of Community Informatics 12(2). https://doi.org/10.15353/joci.v12i2.3246

Flores, M. A. A. (2020). Creating public value with open government data in Latin America. Master's thesis. Westfälische Wilhelms-Universität, Münster.

Germano, E. C., Souza, C. A. de, Sun, V. (2016). Business models adopted by intermediaries in the use of open government data. REBRAE, 9(1): 79–94. https://doi.org/10.7213/rebrae.09.001.AO05

González-Zapata, F., Heeks, R. (2015). Understanding Multiple Roles of Intermediaries in Open Government Data. 13th International Conference on Social Implications of Computers in Developing Countries, May 20, 2015, Negombo, Sri Lanka.

Janssen, M., Charalabidis, Y., Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. Information Systems Management 29(4): 258–268. https://doi.org/10.1080/10580530.2012.716740

Janssen, M., Zuiderwijk, A. (2014). Infomediary Business Models for Connecting Open Data Providers and Users. Social Science Computer Review 32(5): 694–711. https://doi.org/10.1177/0894439314525902

Johnson, P. A., Sieber, R., Scassa, T., Stephens, M., Robinson, P. (2017). The Cost(s) of Geospatial Open Data. Transactions in GIS 21(3): 434–445. https://doi.org/10.1111/tgis.12283

Kitsios, F., Kamariotou, M., Grigoroudis, E. (2021). Digital Entrepreneurship Services Evolution: Analysis of Quadruple and Quintuple Helix Innovation Models for Open Data Ecosystems. Sustainability 13(21): 12183. https://doi.org/10.3390/su132112183

Magalhaes, G., Roseira, C., Manley, L. (2014). Business models for open government data. Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance, 365–370. https://doi.org/10.1145/2691195.2691273

Mayer-Schönberger, V., Zappia, Z. (2011). Participation and power: Intermediaries of open data. 1st Berlin Symposium on Internet and Society, Berlin, Germany.

Osterwalder, A. (2004). The business model ontology a proposition in a design science approach. Doctoral thesis. University of Lausanne.

Osterwalder, A., Pigneur, Y. (2010). Business model generation: A handbook for visionaries, game changers, and challengers (Vol. 1). John Wiley & Sons.

Reggi, L., Dawes, S. (2016). Open Government Data Ecosystems: Linking Transparency for Innovation with Transparency for Participation and Accountability. In H. J. Scholl, O. Glassey, M. Janssen, B. Klievink, I. Lindgren, P. Parycek, E. Tambouris, M. A. Wimmer, T. Janowski, D. Sá Soares (Eds.), Electronic Government (pp. 74–86). Springer International Publishing. https://doi.org/10.1007/978-3-319-44421-5_6

Uhlir, P. F., Schröder, P. (2007). Open Data for Global Science. Data Science Journal 6, 36–53. https://doi.org/10.2481/dsj.6.OD36

van Loenen, B., Zuiderwijk, A., Vancauwenberghe, G., Lopez-Pellicer, F. J., Mulder, I., Alexopoulos, C., Magnussen, R., Saddiqa, M., Rosnay, M. D. de, Crompvoets, J., Polini, A., Re, B., Flores, C. C. (2021). Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. JeDEM - EJournal of EDemocracy and Open Government 13(2): 1–27. https://doi.org/10.29379/jedem.v13i2.644

Voigt, K.-I., Buliga, O., Michl, K. (2017). Business Model Pioneers. Springer International Publishing. https://doi.org/10.1007/978-3-319-38845-8

Xiao, Y., Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. Journal of Planning Education and Research 39(1): 93–112. https://doi.org/10.1177/0739456X17723971

Zhu, Z., Wulder, M. A., Roy, D. P., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., Healey, S. P., Schaaf, C., Hostert, P., Strobl, P., Pekel, J.-F., Lymburner, L., Pahlevan, N., Scambos, T. A. (2019). Benefits of the free and open Landsat data policy. Remote Sensing of Environment 224: 382–385. https://doi.org/10.1016/j.rse.2019.02.016

OPEN GOVERNMENT DATA AND NFTs. A PROPOSED SOLUTION FOR GOVERNMENTS

Michalis Avgerinos Loutsaris[1], **Maria Ioanna Maratsi[1]**, Zoi Lachana[1*], Mohsan Ali[1], Charalampos Alexopoulos[1], Yannis Charalabidis[1]

[1] University of the Aegean, Greece

*correspondence E-mail: zoi@aegean.gr

Keywords: Blockchain Technology (BCT); distributed ledger technology; open data; open government data (OGD); social good; linked data

1. Introduction

Governments, like any other organization, have to provide services based on the data produced through administrative tasks. Governments are responsible for making the interactions with local citizens meaningful in the sense of services. This interaction between governments and citizens can be expensive, but at the same time, it should be two-way communication. Two-way communication, nowadays, is only possible with the help of the necessary tools and technologies required to develop eGovernment strategies, policies, and management. Open Government Data (OGD)-based services and products play a role in improving the living standards of citizens. Since the government is not an isolated structure, open government data is highly dynamic and involves stakeholders such as businesses, citizens, and research sectors. Thus, this complex data flow/exchange and, therefore, interoperability among OGD ecosystems is a challenging task (Charalabidis et al., 2016). An important thing that should be considered is the secure, reliable, consistent, and timely movement of data from one entity to another entity. The quadruple helix can be improved by introducing BCT for secure OGD communication, as shown in **Figure 1**. Blockchain technology is an essential aspect of OGD's transparency and accountability by providing secure blocks with a chaining structure. It is possible to employ openness and transparency to improve public service and budget utilization, increase citizen participation, counteract corruption and fraud, and promote public trust in government.

At the same time, BCT appears to be a tendency to include and utilize this technology beyond the scope of cryptocurrency. Although NFTs are best known for being used to produce digital works of art or collector's items, large companies and public administrations have taken advantage of their potential to make other services or products available to the public: 'analog' items such as the deeds to a house, a city's historical heritage or a company's archive, and smart contracts are some examples of this.

The aim of this research is twofold: On one hand to examine the current state of the art regarding BCT and NFTs in the public sector along, with their potential benefits and barriers and on the other hand to combine NFTs and Open Government Data, by proposing an NFT transactions solution for the public sector.

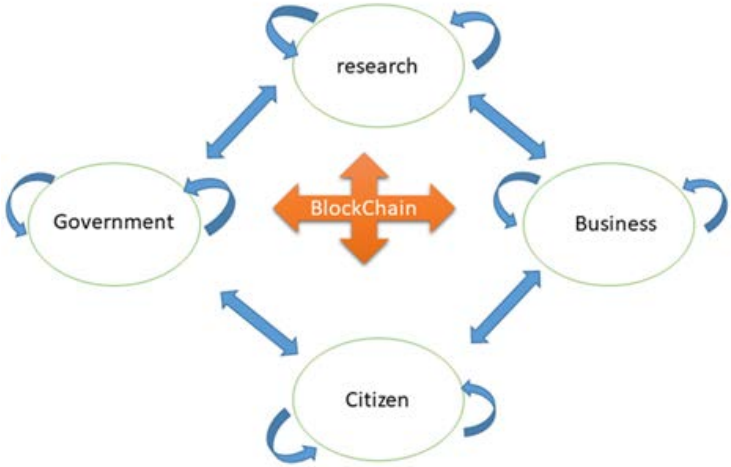

Figure 1. Quadruple helix and BCT-based OD Ecosystem

2. Open Government Data, Blockchain Technology & NFTs

Open Data (OD) is freely available to use and distribute even though for commercial use. There are several other detailed definitions for OD. One of the most appropriate considerations to take into account is the steps needed to be taken from gathering data to completely opening them (**Figure 2**). OD should be complete, fully uploaded, timely for decision-making, primary focus, and mainly collected by experts or researchers. The accessibility refers to their availability on some platforms through APIs or other 5-star formats. Machine readability is an important aspect to consider, although when data is available in csv or xlsx formats, it is considered machine-readable. The data must be non-discriminatory without considering gender or any other factor while opening. Non-proprietary means it must be available under open licensing, which means it can be used and distributed by anyone. Free means that data must be available under the two most commonly used licenses, such as CC-BY and CCO, while publishing the open data. The data also needs to be reviewable to remove outliers, errors, and omissions (E&O) when detected by any end user.

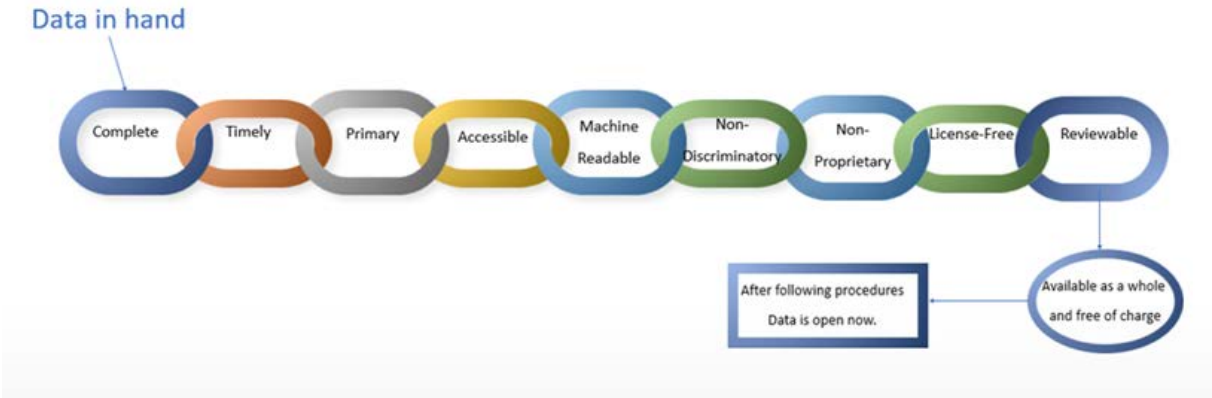

Figure 2. Detailed open data definition

There are three steps/stages involved in the OD plan (**Figure 3**) which are (1) Preparation of OD for publication (2) Launching the OD stage and (3) Extend and sustaining the OD. The preparation stage is a well-defined process where the readiness assessment of the open data is checked by evaluating the technology skills, senior leadership involvement, civic engagement, financing for

OD, data management plan, data demand, policy and legal perspective of OD, and institutions' involvement in OD. After checking their readiness assessment, OD providers build the OD inventory where a mapping between supply and demand is drawn. The legal and policy work is also considered at this stage, such as OD licenses, policies, and legal review. The essential license is selected based on the OD properties and organization rules. The steering committee can be involved to decide better which legal and policy work needs to be more focused on. The last stage refers to the establishment of a data catalog to share the OD with users and the community by means of data curators.

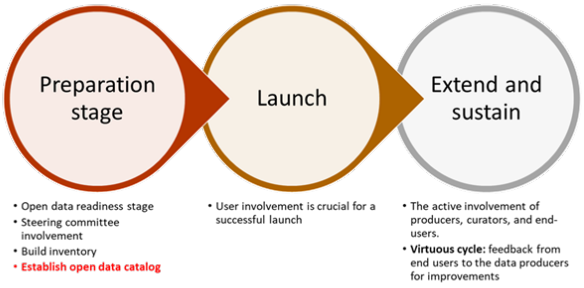

Figure 3. Open data action plan

At all administrative levels, the public sector is one of the largest producers and holders of data and information. In recent years, the size and variety of open data released by public administrations around the world has increased significantly, to the extent that their management is now determined or regulated in many cases, through legislative regulations(Vetrò et al., 2016) Government data is understood as that subset of public data and information produced or collected by the government or government-controlled organizations (Public Agencies and Enterprises). Such data are, for example, applicable laws and circulars, public expenditure, sources of public revenue, etc.(Park et al., 2022). Open government data is essentially a subset of open data and is simply government-related data that is open to the public(Kučera et al., 2013). Several countries have already demonstrated their commitment to open government data by joining the Open Government Partnership(Attard et al., 2015).

Blockchain Technology (BCT, Blockchain as BC) is a piece of software that functions as a ledger distributed across nodes of a communications network. What distinguishes it from other online databases or trading platforms is its immutability since anyone can trade digital assets peer to peer, and no one can alter or undo those transactions without a majority of the network's approval. The main characteristics of BCT, such as its decentralized nature, anonymity of the users, auditability, and others, make it appealing as an infrastructure with great potential but also many challenges. Alexopoulos et al. (2019) conducted a thorough analysis of the benefits and barriers of BCT applications in an eGovernment context and among the key benefits for public services were privacy, data safety, and anonymity (as it is achieved via the usage of private keys). On the other hand, one of the most important challenges was privacy leakage due to the fact that the users' public keys are visible, which can introduce further considerations. On a similar note, Sarantis et al. (2020) investigated BCT in eGovernment and identified fifteen areas in need of related research. Among those areas were data security and privacy, considering identity management and the

protection of personal data with the help of BCT but also under the guidance of concepts such as privacy-by-design and privacy-by-default. Ølnes and Janssen (2018) identified the need to view BCT beyond its currency applications and consider its potential use as an infrastructure or platform for governmental services, such as identity management. They eventually argue that BCT is an emerging platform, as it features decentralized management and control; Ølnes et al. (2017) included the potential benefits and promises of BCT, and under the category "Informational benefits", where, again, data integrity, higher data quality and privacy, can be found. Ølnes and Jansen (2018) developed a framework with the purpose of understanding and making clearer the benefits and challenges associated with the introduction of BCT in public services. Part of this framework was a proof-of-concept example of a project related to the Norwegian Tax Administration, with the results showing that the developed system "worked as expected; however, the immutability of the documents could be problematic with regards to privacy and the GDPR's right-to-be-forgotten". A systematic literature review by Batubara, Ubacht and Jansen (2018) indicated that, however promising, the adoption of BCT for eGovernment lacks empirical evidence and needs further research to target areas such as security, scalability, interoperability, and flexibility. The benefits they see also include improved data integrity, quality, transparency, and privacy; however, they emphasize the need for proper design solutions at an architectural level in order to overcome the identified challenges. Proposed eGovernment systems based on BCT have already been introduced and published, such as the decentralized system based on consortium blockchain technology by Elisa et al. (2019), Alexopoulos et al. (2018).

One type of digital asset is Bitcoin and other cryptocurrencies used in Bitcoin payment networks. Non-fungible tokens are the opposite of these. The term "fungibility" describes the degree to which one asset can be exchanged for another of the same type. Non-fungible assets, in contrast to fungible assets, are valued in a distinct manner due to their specific characteristics and scarcity. Tokens stored on the blockchain that represent non-fungible assets, such as works of art, digital property, or media, are known as NFTs. Whether the item in question is digital or tangible, an NFT can be regarded as a digital certificate of ownership and authenticity that cannot be revoked. There is one unique value represented by each NFT token. No one could reasonably expect to swap out a picture for a different one in a legal agreement without raising suspicions. Antiques or collectibles are just one type of NFT; others include official documents such as birth and death records, property titles, and Internet of Things object identifiers.

The goal of NFT developers was for their creations to be cryptographically verifiable, unique or exceedingly rare, and easy to trade. Using the cryptographic signatures built into the blockchain, the history and present owner of an NFT may be determined in an instant. Because of their adaptability, NFTs can be utilized as substitutes for a wide range of real-world, virtual, and ethereal assets. Digital works of art, digital collectibles, media files, event tickets, smart contracts, and video game assets are the most common NFT assets.

3. The proposed solution

The proposed solution provides transparency and openness on public sector financial movements produced by administrative tasks. In particular, there are three financial movements (or phases):

- The approval decision for a supply or, in general, for some expenditure,

- The commitment decision which commits the budget for the approval decision, and,
- The payment decision after the completeness of payment for the supply or expenditure.

For this purpose, Governments need to build a Government BC in which they will have at least three "Government Wallets", the Main Wallet, the Commitment Wallet, and the Payment Wallet. The Main Wallet is responsible for keeping the approval decisions, the Commitment Wallet is responsible for keeping the commitments decision, and the Payments Wallet is responsible for keeping the payments decision (proofs) after the completeness of the original payment.

The Main Wallet represents the summary of each public organisation's money source, meaning that each public organisation owns its own Main Wallet. At the same time, each VAT number (Citizens, Businesses, etc.) has its own wallet. The Main Wallet and/or each public sector wallet can produce the decisions (Centralized mint), as NFTs, and store the decision in this wallet until the first transaction is made.

During the approval decision, the NFT transferred from the Main Wallet to the Approval Wallet. The same procedure comes for the commitment movement as well. Finally, the payment decision (NFT) is transferred to the VAT number wallet.

The majority of NFTs transactions reside on the Government's cryptocurrency's BC, a distributed public ledger that records transactions. NFTs are individual tokens with valuable information stored in them. The transactions' decisions are stored on the blockchain, but the payments are carried out through bank transfers (same as nowadays). These transactions need to be validated and stored in a block. At the same time, the consensus mechanism for Government BC will be Proof-of-Stake (PoS) to be more energy efficient since Proof-of-Work (PoW) is not energy efficient. Under PoS, Block creators are called validators. A validator checks transactions, verifies activity, votes on outcomes, and maintains records. In return, Validators receive a yearly amount depending on their stake. Under PoW, block creators are called miners. Miners work to solve the hash, a cryptographic number, to verify transactions. In return for solving the hash, they are rewarded with a coin. A Validator is any entity (stakeholder) owning a wallet capable of creating the next block (generate) on the Government's cryptocurrency's blockchain to store the transactions (Decentralized). Validators' wallets are rewarded with government tokens (cryptocurrency).

Government tokens can be redeemed for either discounts on buying goods from green companies that collaborate with governments that foster green sustainable development (electric cars and bicycles) or buying services provided by the public sector (e.g., public gyms, public bicycle rental, etc).

Furthermore, Users can transfer government tokens between their wallets. The redeemed tokens from stakeholders are automatically burned in a burn address. A burn address is a digital wallet that can't be accessed because it doesn't have a private key attached to it, like a lock that someone never built a keyhole for.

The max supply will be infinite, as a pegged coin in order to keep the price as low as it is possible, since the use of this coin is for non-profit reasons. Because of the selected consensus mechanism, at the start of this BC all VAT number wallets will receive 10.000 tokens in order to have the

validation ability. A Validator must have at least 100 tokens in the wallet. Validators can get annual returns between 15% to 30% depending on the Government Tokens they stake. The algorithm may be changed if any new algorithm, less energy efficient, is discovered.

4. Discussion

Eight basic principles are described in the literature (Höchtl and Reichstädter, 2011; Solar et al., 2013), which should meet government data in order to be considered open. These are: (1) Complete; (2) Primary; (3) Timely; (4) Accessible; (5) Machine processable; (6) Non-discriminatory; (7) Non-proprietary; (8) License-free. In parallel, Government data may contain multiple sets of data, including transactions in any form (e.g. financial or not) and expenditure, population, census, parliamentary proceedings, etc. According to Ubaldi (2013) public data sets included in open government data initiatives include:

1.business information (including chamber of commerce information, etc.)
2.registries, patent and trademark information and public databases
3.geographic information (such as address information, aerial photographs, buildings, geology, and topographic information)
4.legal information (such as national, and international court decisions, national laws)
5.meteorological information (including data and models for climate and weather forecasts)
6.social data (such as statistics on the economy, employment, health, population etc.).
7.Transport information (such as traffic congestion, public transport and vehicle classification);

At the same time, the characteristics of NFTs (as it is mentioned above) along with their usage advantages (ownership, authenticity, transferability, creation of economic opportunity, and boosting inclusive growth), constitute them as an innovative solution for governments. The adoption of NFTs solutions by the public sector will not only provide many new capabilities and better services but will establish a safest, openly accessible, transparent and eco-friendly transactional environment. In addition the most important disadvantage of the NFTs usage which is the Concerns Regarding Ecological Impact, seems that fades due to the Proof-of-Stake consensus mechanisms which is proven to be energy efficiency.

Acknowledgements: This research has been financially supported by the European Union under the Marie Skłodowska-Curie ODECO project (grant agreement No. 955569) and the TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Alexopoulos, C., Androutsopoulou, A., Lachana, Z., Loutsaris, M. A., & Charalabidis, Y. (2018). Blockchain technologies in government 3.0: A review. In Proc. EGOV-CeDEM-ePart 11-11.

Alexopoulos, H. et al. (2019), Benefits and Obstacles of Blockchain Applications in e-Government, Proceedings of the 52nd Hawaii International Conference on System Sciences.

Attard, J., Orlandi, F., Scerri, S., Auer, S. (2015). A systematic review of open government data initiatives. Government Information Quarterly 32(4): 399-418. https://doi.org/10.1016/j.giq.2015.07.006

Batubara, R. et al. (2018), Challenges of Blockchain Technology Adoption for e-Government: A Systematic Literature Review, Proceedings of 19th Annual International Conference on Digital Government Research.

Charalabidis, Y., Alexopoulos, C., and Loukis, E. (2016). A taxonomy of open government data research areas and topics. Journal of Organizational Computing and Electronic Commerce 26(1–2): 41–63. https://doi.org/10.1080/10919392.2015.1124720

Elisa, N. et al. (2019), Consortium Blockchain for Security and Privacy-preserving in e-Government Systems, Proceedings of The 19th International Conference on Electronic Business.

Höchtl, J., Reichstädter, P. (2011). Linked open data-a means for public sector information management. In Electronic Government and the Information Systems Perspective 330-343. Springer, Berlin, Heidelberg.

Kučera, J., Chlapek, D., Nečaský, M. (2013). Open government data catalogs: Current approaches and quality perspective. In International conference on electronic government and the information systems perspective 152-166. Springer, Berlin, Heidelberg.

Ølnes, S., Jansen, A. (2017). Blockchain Technology as s Support Infrastructure in e-Government. FIP International Federation for Information Processing.

Ølnes, S. Jansen, A. (2018). Blockchain Technology as Infrastructure in Public Sector – an Analytical Framework. Proceedings of 19th Annual International Conference on Digital Government Research.

Ølnes, S., Ubacht, J., Janssen, M. (2017). Blockchain in government: Benefits and implications of distributed ledger technology for information sharing. Government Information Quarterly 34(3): 355-364. https://doi.org/10.1016/j.giq.2017.09.007

Park, J. J., Kim, E., Jun, H. J. (2022). Development of an Open Government Data (OGD) Evaluation Framework for BIM. Buildings 12(4): 490. https://doi.org/10.3390/buildings12040490

Sarantis, D. et al. (2020), Blockchain in Digital Government: Research Needs Identification, Springer Nature Switzerland.

Solar, M., Meijueiro, L., Daniels, F. (2013). A guide to implement open data in public agencies. In International conference on electronic government 75-86. Springer, Berlin, Heidelberg.

Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. OECD Working Papers on Public Governance, No22, OECD Publishing

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. Government Information Quarterly 33(2): 325-337. https://doi.org/10.1016/j.giq.2016.02.001

OPEN GOVERNMENT DATA (OGD) IN EUROPEAN EDUCATIONAL PROGRAMS CURRICULUM: CURRENT STATE AND PROSPECTS

Nina Rizun[1], Magdalena Ciesielska[1], **Charalampos Alexopoulos[2]**, Stuti Saxena[3*], Georgios Papageorgiou[2]

[1] Gdansk University of Technology, Poland
[2] University of the Aegean, Greece
[3] Graphic Era University, India

*correspondence E-mail: stutisaxenaogd.vishnu@gmail.com

Keywords: Open Government Data; OGD; Universities; Curriculum; Educational Program

It is everybody's knowledge now that Open Government Data (OGD) pertains to the availability of datasets pertaining to government operations and functioning via license-free formats (Afful-Dadzie and Afful-Dadzie, 2017) and these datasets are linked with different themes contingent upon the area of administration like energy, education, climate, tourism, environment, infrastructure, etc. (Ubaldi, 2013). Governments have been benchmarking their OGD initiatives across standard indices like ODIN, OKFN, Open Data Barometer, etc. (Lnenicka et al., 2022; Lnenicka and Nikiforova, 2021) and ample research exists on assessing the quality of OGD portals from the demand (i.e. the governments' efforts at maintaining the quality of datasets) and supply (i.e. the perceptions of users regarding the quality of datasets) side (Crusoe et al., 2019; de Souza et al., 2022; Islam et al., 2021; Kaasenbrood et al., 2015; Khurshid et al., 2022; Lnenicka et al., 2022; Purwanto et al., 2020; Saxena and Janssen, 2017; Shehata and Elgllab, 2021; Talukder et al., 2019; Weerakkody et al., 2017; Wirtz et al., 2016; Wirtz et al., 2018; Wirtz et al., 2019; Zuiderwijk et al., 2015). Given the magnitude of academic interest on OGD- especially in the last 10 years- it remains to be assessed as to how far has the domain progressed in the academic environs and surprisingly, no research has been conducted to elucidate the infusion of this very significant domain- that is relatable to the extent to which the governments are forwarding their claims regarding the furtherance of transparent and corruption-free administration apart from bolstering citizen participation, collaboration and trust (Gil-Garcia et al., 2020; Hellberg and Hedstrom, 2015) besides serving as a means for value creation and innovation by a range of stakeholders (Jetzek et al., 2012; Jetzek et al., 2014) - in the diverse platforms that are meant for furthering academic dialogue and discussion.

Research pertaining to meta-analysis and reviews of the OGD-focused studies already published in academic publication outlets is well-acknowledged (Attard et al., 2015; Kalampokis et al., 2011; Safarov et al., 2017; Wirtz et al., 2022; Wirtz and Birkmeyer, 2015) and to carry forward the baton of OGD-focused research, the present study aims to establish the current baseline of European OGD curricula and describe their fundamental aspects.

The authors applied text analytics and systematic secondary data review methods to explore the nature and scope of existing OGD European study programs to draw a European OGD education current state and prospects. In order to address this aim, this study examines four research questions: R1) What is the current educational offer of the OGD study programs in the top 200 European Universities? R2) What fields of study are the SSC study programs covering? R3) What competencies characterize each cluster of the OGD study programs? R4) What are the common and diverging features of such clusters?

The search for the OGD study programs includes the definition of two types of keyword terms lists and combinations of them. The search combines two types of terms. The first type covered the subject (e.g., OGD). These keywords were searched in the titles and descriptions of the programs. Second, we defined the keywords that characterize training/educational degrees (e.g., MSc). The search combined one term of the subject group and one of the training/educational group (e.g., OGD AND MSc). The terms for search by subject are Open AND Government AND Data, Open AND Data, Open AND Government, Open AND e-Government OR eGovernment), Data AND Management, Government OR e-Government OR eGovernment AND Data, Data AND Reuse, Data AND Analytics AND Government. The terms for search by institutional/educational degree are Bachelor, Capacity Building, Certificate, Continuing Professional Education, Diploma, Education, Executive Master, Graduate, Higher Education, Master, MSc, PhD, Program, Specialization, Training, Undergraduate, Joint Master, MOOC. Google search engine was the main search source. Only English-language search results with geographical location Europe were accepted. Websites of European higher education organizations (public and private) and institutes have been analyzed. Having selected the programs for inclusion in the study, we compiled a database with information derived from their web pages: the program description (program name, the academic level, the aims of the program and/or learning goals, the area of program specialization, competencies, admission requirements); program content (course name, the course type, the course description, the course credits, the learning outcome/goals, course URL); additional information about the institution (name of the institution, the country of the institution, the institution type) and programs (e.g. degree title, the credits-ECTS, the teaching method, the program cost, the program duration, the program URL).

A threefold approach to study the topic was used: (1) text analytics methods (such as text pre-processing, Topic modelling (LDA), Cosine similarity, Elbow and Gap Statistic Methods and the k-means clustering algorithms) to identify OGD educational programs curriculum clusters based on program titles and course description; (2) expert knowledge to review and refine the results against program objectives, descriptions and modules, and also contextual competencies coding and grouping; and (3) statistical analysis to process metadata collected and the results summarization (Ciesielska. et al., 2021).

In total, we identified 21 courses and 5 programs from European Universities and eleven online courses delivered via EdX, Coursera, and MOOCs. As an OGD educational programs curriculum analysis, we identified three clusters: 1) Open Data for Policymakers (30%); 2) Open Data and Digital Transformation (15%); 3) Open and Smart Government (5%); 4) Data Analytics for Government (50%). Each cluster covers specific, non-repetitive content and contains study programs at three levels: supplementary, bachelor and master. We identified 53 unique OGD competencies. The most frequent of them were assigned to the educational programs clusters. The

analysis revealed significant differences in the characteristics of each of identified OGD educational programs clusters, which concern both (i) formal features, such as the use of different teaching methods and program duration, and (ii) conceptual ones, such as lack of consistency in defining roles, responsibilities, competencies and skills to effectively meet the needs of OGD training and imbalances in the interdisciplinary structure of the course/program.

Acknowledgements: The authors acknowledge the financial support from TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Afful-Dadzie, E., Afful-Dadzie, A. (2017). Liberation of public data: Exploring central themes in open government data and freedom of information research. International Journal of Information Management 37(6): 664-672. https://doi.org/10.1016/j.ijinfomgt.2017.05.009

Attard, J., Orlandi, F., Scerri, S., Auer, S. (2015). A systematic review of open government data initiatives. Government Information Quarterly 32(4): 399-418. https://doi.org/10.1016/j.giq.2015.07.006

Ciesielska, M., Rizun, N., Janowski, T. (2021). Interdisciplinarity in Smart Sustainable City education: exploring educational offerings and competencies worldwide. 54th Hawaii International Conference on System Sciences. 2434-2443. https://doi.org/10.24251/HICSS.2021.299

Charalabidis, Y., Alexopoulos, C., Loukis, E. (2016). A taxonomy of open government data research areas and topics. Journal of Organizational Computing and Electronic Commerce 26(1-2): 41-63. https://doi.org/10.1080/10919392.2015.1124720

Crusoe, J., Simonofski, A., Clarinval, A., Gebka, E. (2019). The impact of impediments on open government data use: Insights from users. 13th International Conference on Research Challenges in Information Science (RCIS)1-12. https://doi.org/10.1109/RCIS.2019.8877055

de Souza, A.A.C., d'Angelo, N.J., Filho, R.N.L. (2022). Effects of predictors of citizens' attitudes and intention to use open government data and government 2.0. Government Information Quarterly 39(2): 101663. https://doi.org/10.1016/j.giq.2021.101663

Gil-Garcia, J.R., Gasco-Hernandez, M., Pardo, T.A. (2020). Beyond transparency, participation, and collaboration? A reflection on the dimensions of open government. Public Performance & Management Review 43(3): 483-502. https://doi.org/10.1080/15309576.2020.1734726

Jetzek, T., Avital, M., Bjørn-Andersen, N. (2012). The value of open government data: A strategic analysis framework. 2012 Pre-ICIS Workshop, Orlando, Florida, United States.

Jetzek, T., Avital, M., Bjorn-Andersen, N. (2014). Data-driven innovation through open government data. Journal of Theoretical and Applied Electronic Commerce Research 9(2): 100-120. http://dx.doi.org/10.4067/S0718-18762014000200008

Hellberg, A.S., Hedstrom, K. (2015). The story of the sixth myth of open data and open government. Transforming Government: People, Process and Policy 9(1): 35-51. https://doi.org/10.1108/TG-04-2014-0013

Islam, M.T., Talukder, M.S., Khayer, A. Islam, A.K.M.N. (2021). Exploring continuance usage intention toward open government data technologies: An integrated approach. VINE Journal of Information and Knowledge Management Systems, In press. https://doi.org/10.1108/VJIKMS-10-2020-0195

Kalampokis, E., Tambouris, E., Tarabanis, K. (2011). A classification scheme for open government data: Towards linking decentralised data. International Journal of Web Engineering and Technology 6(3): 266-285.

Kaasenbrood, M., Zuiderwijk, A., Janssen, M., de Jong, M., & Bharosa, N. (2015). Exploring the factors influencing the adoption of open government data by private organizations. International Journal of Public Administration in the Digital Age 2(2): 75-92. https://doi.org/10.4018/ijpada.2015040105

Khurshid, M.M., Zakaria, N.H., Arfeen, M.I., Rashid, A., Nasir, S.U., Shehzad, H.M.F. (2022). Factors influencing citizens' intention to use open government data-A case study of Pakistan. Big Data and Cognitive Computing 6(1): 31. https://doi.org/10.3390/bdcc6010031

Lnenicka, M., Luterek, M., Nikiforova, A. (2022). Benchmarking open data efforts through indices and rankings: Assessing development and contexts of use. Telematics and Informatics 66: 101745. https://doi.org/10.1016/j.tele.2021.101745

Lnenicka, M., Nikiforova, A. (2021). Transparency-by-design: What is the role of open data portals? Telematics and Informatics 61: 101605. https://doi.org/10.1016/j.tele.2021.101605

Lnenicka, M., Nikiforova, A., Saxena, S., Singh, P. (2022). Investigation into the adoption of open government data among students: The behavioral intention-based comparative analysis of three countries. Aslib Journal of Information Management 74(3): 549-567. https://doi.org/10.1108/AJIM-08-2021-0249

Open Data Barometer (2022). Open Data Barometer. Available at https://opendatabarometer.org/?_year=2017&indicator=ODB, accessed on August 17, 2022.

ODIN. (2022). Open Data Inventory. Available at https://odin.opendatawatch.com/, accessed on August 17, 2022.

OKFN. (2022). Open Knowledge Foundation. Available at https://index.okfn.org/, accessed on August 17, 2022.

Purwanto, A., Zuiderwijk, A. Janssen, M. (2020). Citizens' trust in open government data: A quantitative study about the effects of data quality, system quality and service quality. In: The 21st Annual International Conference on Digital Government Research, ACM, 310-318.

Safarov, I., Meijer, A., Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. Information Polity 22(1): 1-24. https://doi.org/10.3233/IP-160012

Saxena, S., Janssen, M. (2017). Examining open government data (OGD) usage in India through UTAUT framework. Foresight 19(4): 421-436. https://doi.org/10.1108/FS-02-2017-0003

Shehata, A., Elgllab, M. (2021). Saudi scholars' perceptions and use of open government data portals at Shaqra University, Saudi Arabia. International Federation of Library Associations and Institutions (IFLA Journals) 47(4): 493-504. https://doi.org/10.1177/03400352211023834

Talukder, M.S., Shen, L., Talukder, M.F.H., Bao, Y. (2019). Determinants of user acceptance and use of open government data (OGD): An empirical investigation in Bangladesh. Technology in Society 56: 147-156. https://doi.org/10.1016/j.techsoc.2018.09.013

Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives. OECD Working Papers on Public Governance, 22. OECD Publishing Press. https://doi.org/10.1787/5k46bj4f03s7-en

Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., Dwivedi, Y.K. (2017). Open data and its usability: An empirical view from the citizen's perspective. Information Systems Frontiers 19: 285-300. https://doi.org/10.1007/s10796-016-9679-1

Wirtz, B.W., Becker, M., Weyerer, J.C. (2022). Open government: Development, concept, and future research directions. International Journal of Public Administration. https://doi.org/10.1080/01900692.2021.2019273

Wirtz, B.W., Birkmeyer, S. (2015). Open government: Origin, development, and conceptual perspectives. International Journal of Public Administration 38(5): 381-396. https://doi.org/10.1080/01900692.2014.942735

Wirtz, B.W., Piehler, R., Thomas, M.J., Daiser, P. (2016). Resistance of public personnel to open government: A cognitive theory view of implementation barriers towards open government data. Public Management Review 18(9): 1335-1364. https://doi.org/10.1080/14719037.2015.1103889

Wirtz, B.W., Weyerer, J.C., Rosch, M. (2018). Citizen and open government: An empirical analysis of antecedents of open government data. International Journal of Public Administration 41(4): 308-320. https://doi.org/10.1080/01900692.2016.1263659

Wirtz, B.W., Weyerer, J.C., Rosch, M. (2019). Open government and citizen participation: An empirical analysis of citizen expectancy towards open government data. International Review of Administrative Sciences 85(3): 566-586. https://doi.org/10.1177/0020852317719996

Zuiderwijk, A., Janssen, M., Dwivedi, Y.K. (2015). Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology. Government Information Quarterly 32(4): 429-440. https://doi.org/10.1016/j.giq.2015.09.005

OPEN DATA DEVELOPMENTS, OPPORTUNITIES AND CHALLENGES IN EUROPE: LESSONS LEARNED FROM OPEN DATA BEST PRACTICES IN 2022

Frederika Welle Donker[1*], Ana Kuveždić Divjak[2]

[1] Faculty of Architecture & the Built Environment, Delft University of Technology, Netherlands
[2] Faculty of Geodesy, University of Zagreb, Croatia

*correspondence E-mail: f.m.welledonker@tudelft.nl

Keywords: high value datasets; open data research; capacity building; open data ecosystems; data spaces

1. Introduction to site visits as part of Capacity Building

As part of the Twinning Open Data Operational (TODO) Project, a series of virtual and physical site visits were organized in 2021 and 2022 as part of the Capacity Building Work Package. The goal of the site visits was to increase the knowledge of the TODO Consortium partners and to expand their understanding of the real-life implementation of open data policies, technologies, and initiatives in different domains. The institutions that participated in the site visits are considered to be among open data best practices in Europe. These organizations included public organizations at different administrative levels (EU, national, regional, local), businesses, research institutions and non-profit organizations. The aim of the site visits was to gain insight in policies and day-to-day practices of open data organizations and in the main challenges they are facing. This paper describes the outcomes of these site visits, and which open data developments, challenges and opportunities are currently going on for open data organisations.

In the original planning, two rounds of site visits for research staff of the University of Zagreb (UNIZG) were planned, with the first round in 2020 and the second round in 2021. However, due to COVID-19 travel restrictions, the first round of site visits was initially postponed from 2020 to 2021. When it appeared that the travel restrictions would not be lifted in the short term in 2021, we decided to organise virtual site visits as online events instead of the first round of physical site visits. Although online meetings have their drawbacks, it was deemed preferable to postponing the site visits yet again, especially as other Work Packages of the TODO project relied on these capacity building events. The virtual site visits also had some advantages. Firstly, more people were able to attend the events. The virtual site visits were open to all interested TODO partners, Early-Stage Researcher (ESRs) and other stakeholders. Secondly, we were able to invite more speakers from organisations that would not have been able to host a group of visitors.

2. Details of the site visits

The first series of two online sessions were held in the last week of June and first week of July 2021. The second series of four online knowledge exchange sessions were held in the last quarter of 2021. The online sessions were organized per domain/topic, or per country. These sessions allowed TODO partners to learn about the latest developments in open data policies and implementation, and to exchange experiences. The invited Open Data speakers presented their experiences and in an open discussion talked about not only the positive effects but also which barriers they have encountered, or which challenges are still to be addressed. The first online session was held as a hybrid event with MSc students at Delft University of Technology (TU Delft) with UNIZG staff attending online. All other sessions were held as online only events. The online sessions were attended by an average of 25 attendees, which was roughly double the number of intended participants of the original physical site visit.

As travel restrictions were eased during February 2022, we were able to organise a physical site visit. The site visit was held between 22 and 24 February 2022 in Brussels. Brussels was chosen as most of the actors and organizations active in the different phases of the open data life cycle either have an office in Brussels or were able to travel to Brussels due to its central position in Europe. Unfortunately, due to illnesses and some restrictions still in place, we were not able to visit all the scheduled open data actors and organisations. **Table 1** provides an overview of the dates of the sessions, the focus of the meeting and the participating organisations.

Table 1. Details of the virtual and physical site visits

Date	Focus	Participating organisations	Comments
29 June 2021	Managing (High-Value) Open Data Platforms for geodata and statistical data	• Netherlands' Cadastre, Land Registry and Mapping Agency (Kadaster); • Statistics Netherlands (CBS)	Hybrid event at TU Delft with MSc. students, TODO Consortium members and ESRs
6 July 2021	Balancing public sector open data and private sector non-open data	• NDW – Netherlands National Road Traffic Data Portal; • Open State Foundation (Dutch NGO advocating open data)	Online event attended by TODO Consortium members and ESRs
15 Nov. 2021	Legal aspects of open (research) data and balancing data protection	• Dr. Alexandra Giannopoulou, Institute for Information Law (IViR), University of Amsterdam; • Dr. Lorenzo Della Corte, University of Tilburg	Online event attended by TODO Consortium members and ESRs
22 Nov. 2021	Data Spaces and Urban Data Platforms	• Renata Ávila of the Open Knowledge Foundation Network (OKF);	Online event attended by TODO Consortium members and ESRs

		• Dr. Marcel van Oosterhout of Rotterdam School of Business	
29 Nov. 2021	Implementing a National Data Strategy in the United Kingdom	• UK Department for Digital, Culture, Media & Sport (DCMS); • Transparency & Data Ethics Team, Central Digital and Data Office, Cabinet Office; • Open Data Institute UK	Online event attended by TODO Consortium members and ESRs
6 Dec. 2021	Open Data in Estonia	• Prof. Robert Krimmer, Tartu University; • Government CIO Office, Estonia; • Dr. Tomislav Vracić, Microsoft	Online event attended by TODO Consortium members and ESRs
8 Feb. 2022	European Data Protection – present and future	• Dr. Stefano Leucci, European Data Protection Supervisor	Online event attended by TODO Consortium members and ESRs
23 Feb. 2022	Open Geodata – opportunities and challenges	• National Geographe Institut (NGI); • EuroGeographics • PSI Alliance	Physical meeting at Royal Military School attended by TODO consortium members and ESRs
24 Feb. 2022	Open Data Implementation in Flanders	• Informatie Vlaanderen; • Open Knowledge Belgium	Physical meeting at the Government of Flanders building, attended by TODO Consortium members and ESRs

3. Main lessons learned from the site visits

From all the sessions, it emerged that although many steps have already been taken to improve accessibility of government open data in many European countries, there are still challenges to be tackled, even by the open data leaders, such as the United Kingdom, The Netherlands and Estonia.

For open data providers one of the main challenges is how to get in touch with open data (re)users in an open data ecosystem. It is important that data providers and data users are in contact to ensure that their specific open data needs are met, such as which data formats and with which specifications. Many open data portals do not require user registration, therefore, there is no way to engage directly with (re)users. Open data providers need to explore other ways to stay in contact with open data users.

Another challenge that was often reflected upon, is the lack of resources in terms of human resources, and financial resources. Both data providers and open data users need more technically skilled personnel to keep up with emerging technologies and ensuing issues, such as data ethics and personal data protection. Open data portals and platform managers need an ongoing financial

commitment to keep the services running, especially if there is an ambition or a necessity to provide a data platform with interactive services and 24/7 APIs, rather than just providing static data catalogue services. With a rapid increase in the amount of data, storage capacity requires more resources, especially for storing historical datasets and older versions after the periodical update. Such datasets can be particularly useful for trend analyses. However, is it within the mandate of the public sector to keep these datasets available as open data, especially if the public sector body concerned is a self-funding agency, responsible for generating income to recover a substantial part of their operating costs?

Trust in general was mentioned in almost all sessions. Trust of citizens and companies in the public sector to provide data in a findable way; trust in the veracity of the data, trust in the way decisions are made; especially when new technologies such as algorithms and artificial intelligence (AI) are employed; and trust that processes are transparent and accountable. And vice versa, trust of the public sector in the private sector that data are reused for beneficial purposes and not misused. And finally, both citizens and the public sector need trust in companies participating in public-private projects, such as Smart City projects. One option for sharing data between the different stakeholders is via data spaces. This concept needs to be researched further.

A lack of digital skills both within the public sector as well as within society is viewed as another major challenge. With rapid technological changes, such as AI, and an increasing use of algorithms, it becomes harder for citizens to understand in which way governments and companies deal with data. But it also becomes harder for the public sector to be able to cope with the ever-increasing amount of data and having to develop new e-services to comply with digital transition strategies. Low levels of digital maturity prove to be a major barrier to building public trust. And without public trust, an open data ecosystem cannot exist, especially considering upcoming developments such as the European Commission's proposals for the Data Act and the Data Governance Act, and digital transformation strategies. Estonia shows that teaching digital skills should start at kindergarten level to ensure a sufficient level of digital maturity. However, keeping digitally skilled citizens in the country is another challenge, as many skilled people migrate to another country with more opportunities.

Finally, recent innovative technologies, such as AI, create new challenges, such as data ethics and (re)assessment of how to deal with the balance between open data and protection of personal data.

Acknowledgments: This research is part of the Twinning Open Data Operational project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 857592.

SESSION II

OPEN DATA ENHANCING TRANSPARENCY

HOW DO OGD INITIATIVES AFFECT TRANSPARENCY? A POST-ANALYSIS ON DEVELOPED INDICES

Charalampos Alexopoulos[1], **Nina Rizun[2]**, Stuti Saxena[3*]

[1] University of the Aegean, Greece
[2] Gdansk University, Poland
[3] Graphic Era University, India

*correspondence E-mail: stutisaxenaogd.vishnu@gmail.com

Keywords: Open Government Data; OGD; Open Data Inventory; ODIN; Transparency; Corruption Perception Index; CPI; World Press Freedom Index; WPFI; Quadrant analysis

Open Government Data (OGD) refers to the datasets pertaining to the structural, functional and operational dimensions of government bodies and are available freely (in most cases) via dedicated web portals in machine-processable formats with government-mandated open license. OGD is available across multifarious domains like climate, energy, education, transport, finance, and a humungous set of sectors where the government operates (Janssen, 2011). As an advanced stage of e-government (electronic government) where the impetus is on citizen collaboration and participation in administration, OGD initiatives seek to refurbish public service delivery with this novel administrative innovation (Vetro et al., 2016). The launch of OGD initiatives by the governments around the globe is a phenomenal action in several countries. OGD is re-used by diverse set of stakeholders (academia, professionals, businesses, non-governmental organizations, and the like) for value creation and innovation. OGD is also known to bring about economic growth (Attard et al., 2015). Nonetheless, one of the main goals of any OGD initiative is to improve transparency of and trust on governmental operations and decisions (Gonzalvez-Gallego et al., 2020).

Extant literature on OGD has advanced a number of conceptual frameworks to further our understanding of the domain. Various studies have investigated the users' intention to use, adopt and acknowledge the utilitarian facet of OGD in terms of product and/or service improvisation (Dawes et al., 2016; Sieber and Johnson, 2015; Srimuang et al., 2017). Fewer concentrate on the transparency factors influenced by OGD or vice versa (Abu Shanab, 2015a; Abu-Shanab, 2015b; Jetzek et al., 2013; Lee and Kwak, 2011; Sandoval-Almazan and Gil-Garcia, 2016; Wirtz and Birkmeyer, 2015). The latter ones argue that there is a connection between the impact of GDP or GDP growth rate with the implementation of OGD initiatives and, ipso facto, with the rate of participation and trust. But most of the above studies are either qualitative or conceptual based on experts' opinions. Thus, they are not based on absolute quantitative metrics and indices. This study tries to fill in this gap by investigating the correlation and the degree of influence of open data on transparency, GDP and trust using well-established indicators. Thus, the present study addresses the call made by Charalabidis and his colleagues that OGD research is required to appreciate the

"use of OGD as part of anticorruption programs in order to increase public sector accountability and credibility" (Charalabidis et al., 2016).

Across the globe, countries have been taking steps to curb corruption and bring about transparency in their functioning. Furthermore, governments are resorting to bolster trust among the citizens regarding the efficacy of the government functioning. Connecting the dots with the foregoing, countries which have spearheaded the OGD initiatives already are also stepping up their drives towards ensuring that their OGD portals are as much "open" as possible, wherein openness is defined in terms of transparency, completeness, accuracy, timeliness, and other key criteria. Capturing these aforesaid dimensions in the present study along with the dimension of economic growth of the countries is the key contribution of the present study. Specifically, whereas, there has been an impetus on countries' aiming to improve their performance across the transparency rankings (Corruption Perception Index as per the Transparency International rankings or the World Press Freedom Index (WPFI) as per the Reporters Without Frontiers) and the GDP (Gross Domestic Product) indices, concomitantly, governments have been vying top-notch rankings on the ODIN (Open Data Inventory that measures the degree of openness of the OGD portals) indicator as well. Furthermore, governments are striving to forge ties with the citizens to bolster the trust of the latter in them (GovData360).

Corruption Perception Index (CPI)-released by Transparency International- is a widely-acclaimed indicator against which countries are ranked on the basis of the perceptions regarding corruption as far as the public sector is concerned (Transparency International, 2022). CPI assesses dimensions such as bribery, diversion of public funds, public officials' making private gains, red-tapism, nepotism, etc. World Press Freedom Index (WPFI) is annually released by the Reporters Without Frontiers wherein the freedom of information and communication is gauged (RSF, 2022). The Gross Domestic Product (GDP) helps to assess the "health" of a country's economy such that the value of all the goods and services produced over a particular time period is specified-this is also referred to as the size of the economy (Talukdar, 2013). Trust in government may be captured by the indices referred to in the indices like citizens' participatory competence, civil rights, constitutional discretion, etc.(GovData360, 2022). Open Data Inventory (ODIN) evaluates the extent to which the open data provided via the online portal are "open" and "complete" on the bases of machine readability (allow users to easily process data using a computer, e.g. XLS, XLSX, CSV, Stata, SAS, SPSS, JSON, CDF, RDF, XML, and TXT files), non-proprietary format (allow users to access data without requiring the use of a costly, proprietary software, e.g. PDF, HTML, XLSX, DOCX, CSV, JSON, XML, and TXT files), download options (bulk download, API, and user-select options), metadata availability (including definition of the indicator, date of upload to website or when dataset was last updated and name of data source) and terms of use (with "open" terms of use or open data license) (ODIN, 2022).

The purport of the study is to situate a country's OGD initiative across four axes-CPI, WPFI, GDP and GovData360 rankings vis-a-vis ODIN. There are three reasons for selecting these 4 metrics: first, OGD initiative is an attempt to bolster transparency and the same gets reflected in the CPI which measures the extent of corruption whilst WPFI captures the freedom of speech and expression, and hence transparency; second, GDP a country's GDP indicates its economic well-being and to launch and institutionalize an OGD initiative, a country's GDP should be factored into; and, finally, ODIN assesses the extent to which the datasets available on the country's OGD

portal meet the internationally-acknowledged "openness" metrics, and, hence reflective of the quality of the datasets which, in turn, is symbolic of the extent of progressive and advanced state of an OGD initiative. In line with these dimensions, a country's OGD initiative is explained using a conceptual framework. Whereas studies abound regarding the conceptual frameworks linked with OGD, an understanding of the countries' status vis-a-vis the parameters chosen in this study merits consideration-the present study seeks to plug the gap.

That there is a linkage between corruption (as measured by CPI and WPFI) and economic development (as measured by GDP) has been attested (Tanzi and Davoodi, 1998; Ibraheem et al., 2013; Lucic et al., 2016; Grundler and Potrafke, 2019). Likewise, extant research on OGD shows that OGD initiatives and GDP are related with each other (Magalhaes et al., 2013; Saxena, 2017; Leviakangas and Molarius, 2020; Zeleti et al., 2016; Gonzalez-Zapata and Heeks, 2015) -however, empirical investigation on this linkage has remained few and far between. Finally, there has been scant research to study the linkage between the ODIN rankings and transparency (which is linked with trust and, inter alia, corruption) (See Gonzalvez-Gallego and Nieto-Torrejon, 2021). Therefore, it is justified to appreciate the manner in which the three dimensions are inter-linked in terms of the OGD initiatives of the governments. Specifically, we will plot the countries in the quadrants (**Table 1**).

Table 1. Quadrant framework to understand the OGD initiative across Transparency ranking-GDP axes-ODIN scores (Source: Authors).

	GDP (High)	**GDP (Low)**
Transparency ranking/ODIN score (High)	Established OGD initiative (Eg: The Netherlands)	Work-in-progress OGD initiative (Eg: Greece; Spain)
Transparency ranking/ODIN score (Low)	Crawling OGD initiative (Eg: India)	Shelve-it-off OGD initiative (Eg: Bhutan)

Backed by data on 28 countries, the present study seeks to present a quadrant analysis validated with an empirical framework wherein the correlations between the transparency, trust and GDP rankings shall be studied with the lens of the ODIN rankings. These correlation-based deductions shall be followed up by a regression analysis wherein the ODIN serves as the dependent variable and the GDP figures, Trust indices and CPI indices shall figure as independent variables. Findings from the data analyses pitch, for instance, The Netherlands in the "Established", Greece in the "Work-in-progress", India in the "Crawling" and Sri Lanka in the "Shelve-it-off" quadrants. Likewise, other countries have been mapped across the axes to drive home the linkage between the indices.

This study makes a significant contribution towards furthering our understanding of the OGD phenomenon regarding its correlation with transparency, trust and GDP. The results indicate strong correlations among all these factors providing a robust empirical validation on the existing theoretical and conceptual background. Moreover, the study presents the developed clusters of OGD initiatives and discusses them regarding the Open Data Maturity Index. Future research pointers and practitioner implications constitute the concluding segments of the study.

Acknowledgements: The authors acknowledge the financial support from TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Abu-Shanab, E. A. (2015a). Open government initiatives in public sector: A proposed framework for future research. Saba Journal of Information Technology and Networking 3(1): 4–14.

Abu-Shanab, E. A. (2015b). Reengineering the open government concept: An empirical support for a proposed model. Government Information Quarterly 32(4): 453-463. https://doi.org/10.1016/j.giq.2015.07.002

Attard, J., Orlandi, F., Scerri, S., Auer, S. (2015). A systematic review of open government data initiatives. Government Information Quarterly 32(4): 399-418. https://doi.org/10.1016/j.giq.2015.07.006

Charalabidis, Y., Alexopoulos, C., Loukis, E. (2019). A taxonomy of open government data research areas and topics. Journal of Organizational Computing and Electronic Commerce 26(1-2): 41-63. https://doi.org/10.1080/10919392.2015.1124720

Dawes, S.S., Vidiasova, L., Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. Government Information Quarterly 33(1): 15-27. https://doi.org/10.1016/j.giq.2016.01.003

Gonzalez-Zapata, F., Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. Government Information Quarterly 32(4): 441-452. https://doi.org/10.1016/j.giq.2015.09.001

Gonzalvez-Gallego, N., Nieto-Torrejon, L. (2021). Government data openness and coverage. How do they affect trust in European countries? Journal of Data and Information Science 6(1): 139-153. https://doi.org/10.2478/jdis-2021-0010

Gonzalvez-Gallego, N., Nieto-Torrejon, L., Perez-Carceles, M.C. (2020). Is open data an enabler for trust? Exploring the link and the mediating role of citizen satisfaction. International Journal of Public Administration 43(14): 1218-1227. https://doi.org/10.1080/01900692.2019.1668412

GovData360 (2022). World Bank. Available at https://govdata360.worldbank.org/indicators/. Accessed on May 30, 2022.

Grundler, K., Potrafke, N. (2019). Corruption and economic growth: New empirical evidence. European Journal of Political Economy 60, 101810. https://doi.org/10.1016/j.ejpoleco.2019.08.001

Ibraheem, N.K., Umar, G., Ajoke, A.F. (2013). Corruption and economic development: Evidence from Nigeria. Kuwait Chapter of Arabian Journal of Business and Management Review 3(2): 46-56.

Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. Government Information Quarterly 28(4): 446-456. https://doi.org/10.1016/j.giq.2011.01.004

Jetzek, T., Avital, M. Bjorn-Andersen, N. (2013). Generating value from open government data. Thirty Fourth International Conference on Information Systems, Milan.

Lee, G., Kwak, Y.H. (2011). Open government implementation model: A stage model for achieving increased public engagement. In Proceedings of the 12th Annual International Conference on Digital Government Research. https://dl.acm.org/doi/pdf/10.1145/2037556.2037598

Leviakangas, P., Molarius, R. (2020). Open government data policy and value added- Evidence on transport safety agency case. Technology in Society 63: 101389. https://doi.org/10.1016/j.techsoc.2020.101389

Lucic, D., Radisic, M., Dobomirov, D. (2016). Causality between corruption and the level of GDP. Economic Research-Ekonomska Istraživanja 29(1): 360-379. https://doi.org/10.1080/1331677X.2016.1169701

Magalhaes, G., Roseira, C., Strover, S. (2013). Open government data intermediaries: A terminology framework. Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance 330-333. https://doi.org/10.1145/2591888.2591947

ODIN (2022). Open Data Inventory. Available at https://odin.opendatawatch.com/Report/faq. Accessed on May 23, 2022.

RSF (2022). Reporters Without Frontiers. Available at https://rsf.org/en/index. Accessed on September 3, 2022.

Sandoval-Almazan, R. Gil-Garcia, J.R. (2016). Toward an integrative assessment of open government: Proposing conceptual lenses and practical components. Journal of Organizational Computing and Electronic Commerce 26(1-2): 170-192. https://doi.org/10.1080/10919392.2015.1125190

Saxena, S. (2017). Prospects of open government data (OGD) in facilitating the economic diversification of GCC region. Information and Learning Sciences 118(5-6): 214-234. https://doi.org/10.1108/ILS-04-2017-0023

Sieber, R.E., Johnson, P.A. (2015). Civic open data at a crossroads: Dominant models and current challenges. Government Information Quarterly 32(3): 308-315. https://doi.org/10.1016/j.giq.2015.05.003

Srimuang, C., Cooharojananone, N., Tanlamai, U., Chandrachai, A. (2017). Open government data assessment model: An indicator development in Thailand. ICACT. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7890110

Talukdar, M.H. (2013). Economy and transparency: The model invention. Economia. Seria Management 16(2): 286-296.

Tanzi, V., Davoodi, H. (1998). Corruption, public investment, and growth. In: Shibata, H., Ihori, T. (eds.) The Welfare State, Public Investment, and Growth. Springer, Tokyo.

Transparency International (2022). The ABCs of the CPI: How the corruption perceptions index is calculated. Transparency International. Available at https://www.transparency.org/en/news/how-cpi-scores-are-calculated. Accessed on May 21, 2022.

Vetro, A., Canova, L., Torchiano, M., Minotas, C.O., Lemma, R., Morando, F. (2016). Open data quality measurement framework: Definition and application to open government data. Government Information Quarterly 33(2): 325-337. https://doi.org/10.1016/j.giq.2016.02.001

Wirtz, B.W. Birkmeyer, S. (2015). Open government: Origin, development, and conceptual perspectives. International Journal of Public Administration 38(5): 381-396. https://doi.org/10.1080/01900692.2014.942735

Zeleti, F.A., Ojo, A., Curry, E. (2016). Exploring the economic value of open government data. Government Information Quarterly 33(3): 535-551. https://doi.org/10.1016/j.giq.2016.01.008

ANALYZING THE IMPACTS OF OPEN DATA PORTALS ON URBAN DEVELOPMENT AND INNOVATIONS IN GERMAN CITIES

Mohit Kapoor[1*], Mathias Jehling[1]

[1] Leibniz Institute of Ecological Urban and Regional Development (IOER), Dresden, Germany

*correspondence E-mail: m.kapoor@ioer.de

Keywords: Open data; Urban innovation; Evidence-based planning; Data-driven decision-making; Data portal

1. Introduction

In the last ten years, thanks to technological innovations, there has been a fundamental change in urban planning and governance. The availability of data sets, IT solutions or online dashboards has enabled cities to use them for efficient public services, thereby also improving the quality of life in cities. One of the most important approaches here is the use of open data. Open data is seen as central to enabling other services and innovations for society in addition to improving processes within public administration (Horáková 2017; Linåker, Runeson 2020). Open data portals play a major role here (Publications Office of the European Union 2022). However, the increasing opportunities for urban development also raise questions regarding implementation at the municipal level.

The aim of the article is to use the example of Berlin and Dresden to learn more about the functionality of open data portals from the user's perspective and to present the barriers and pathfinders for the accessibility, usability and impact of open data portals at the municipal level. From the point of view of spatial planning and urban development, the advantages for public participation, transparency of administrative actions and fair urban development should be worked out. The central research questions in this regard are a) How do different stakeholders use open data portals in the city of Berlin and Dresden? and b) How effective are they for data-driven planning and decision-making?

The use of open data for planning and communication between city politics, administration, citizens and civil society is the focus of this study. According to Tim-Berners Lee 5-Star model, cities have the opportunity to increasingly use this data for spatial planning, especially in the areas of land use, housing, traffic, infrastructure, environment and Communication. Key reasons for using open data for spatial planning and development are:

a) Transparency - in order to democratize planning procedures and monitor administrative action. To achieve this, it is important that citizens and other stakeholders not only have easy access to the data, but can also use it freely and effectively.

b) social and economic added value - for any dataset such as geospatial, traffic, health or environmental data, there may be a number of other purposes beyond what was initially envisaged. The creative freedom to innovate with these data sets makes it possible to better exploit their social and economic value.

c) Participatory governance - because through open data initiatives, citizens do not remain passive "recipients" of urban development, but can get involved as active stakeholders. The cooperation between citizens and administration is better structured and more regular, which improves participation in the cities (Attard et al., 2015).

2. Methodology

For this study, Berlin and Dresden was used as a case study as it has been pursuing an open data strategy since 2011 and 2016 respectively. Both the open data portals in both cities have a comprehensive range of multi-format data sets and a wide range of applications. In addition, there is strict coordination of the provision of open data within the administration.

In order to analyze the availability and impact of these two open data portals on urban development, twelve expert interviews were conducted in Berlin's open data ecosystem and ten expert interviews in Dresden. The selected experts hold important positions and have published studies, some on open data. All interviews were held between April and August 2022, and included representatives from urban planning, IT, data science, the Senate, business, research and civil society. During the interviews, respondents were asked qualitative questions about the accessibility and usability of open data portals and their impact on urban development, planning, governance and aspects of city management. The interviews were recorded and then transcribed.

Comparable methods of content analysis and frameworks by (Kuckartz and Rädiker 2022; Máchová et al. 2018; Norton, 2008) were used to derive a user-centered evaluation approach with the help of content analysis. The transcripts were coded to derive categories that can be used to identify key barriers and enablers. Through the interviews, we classified barriers and enablers in terms of a) accessibility – the ease of finding relevant datasets, b) usability – the quality and interoperability of the datasets, and c) impact – the desired outcomes and improvements for urban development (Nikiforova, 2022).

3. Results and Way Forward

Berlin and Dresden are at different open-data maturity levels, hence barriers and enablers vary: Berlin completes 11 years in its Open Data journey and is developing Open Data Strategy 2.0 with an aim to achieve 5-Star Linked Open Data Model, strengthen data governance and public relations for Civic Tech (**Figure 1**). Dresden completes 6 years and requires a robust long-term strategy to address the present disintegrated approach, with a further aim to strengthen open data ecosystem promoting its use for urban innovations (**Figure 2**).

Large Cities (5% of Germany) vs Medium and Small Towns/Cities (95% of Germany): The economic capital, access to technological tools and human resource expertise available to large cities makes them frontrunners to benefit from open data for improved urban planning, provision

of urban services and participatory governance activities. Medium and Small towns/cities face challenges to raise economic resources and human capital to adopt digital tools for urban governance and innovations. Hence, to address this disparity, special emphasis in the form of Open Data Strategy/Policy, annual budgets, human resources, technical expertise and capacity building are required to empower such towns/cities.

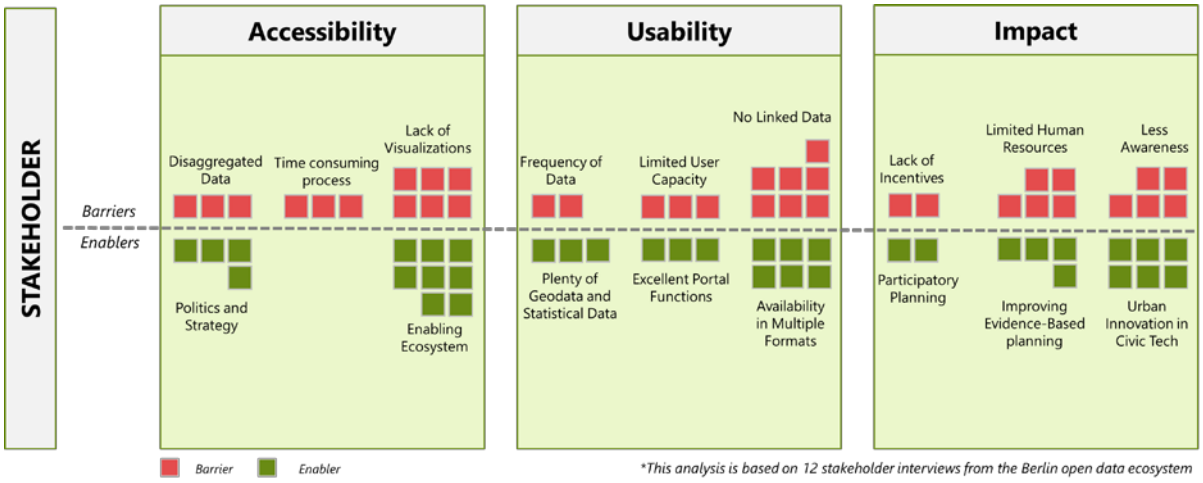

Figure 1. Results from Berlin

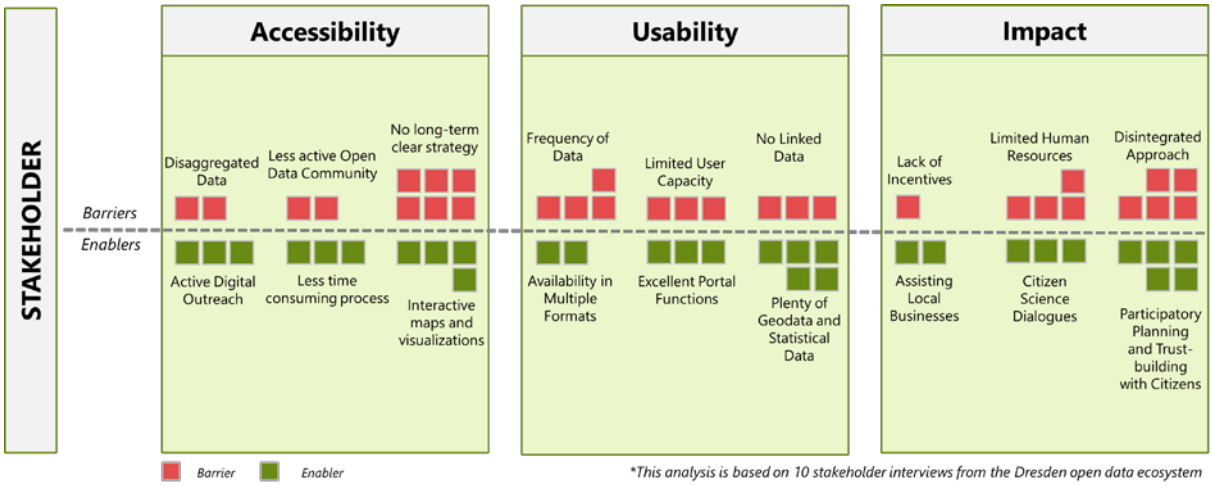

Figure 2. Results from Dresden

The use of open data in spatial planning is in its initial stages and the trend towards increased adoption of digitization with more updated data sets, including live data and big data, has the potential to become 'mainstream' in urban development activities.

References

Attard, J., Orlandi, F., Scerri, S., Auer, S. (2015). A systematic review of open government data initiatives. Government Information Quarterly 32(4): 399–418. https://doi.org/10.1016/j.giq.2015.07.006

Both, W., Schieferdecker, I. (Hrsg.) (2012). Executive Summary Berlin Open Data Strategy - Organisational, legal and technical aspects of Open Data in Berlin, Concept, pilot system and recommendations for action.

Bürger, T., Hoch, A., Bertelsmann Stiftung (2020). Open Data in Kommunen: Eine Kommunalbefragung zu Chancen und Herausforderungen der Bereitstellung offener Daten. Leb. Kommune 4. https://doi.org/10.11586/2020068

Colpaert, P., Joye, S., Mechant, P., Mannens, E., de Walle, R.V. (2013). The 5 stars of open data portals. Proceedings MeTTeG 2013(7) 61-67.

de Juana-Espinosa, S., Luján-Mora, S. (2020). Open government data portals in the European Union: A dataset from 2015 to 2017. Data Brief 29: 105156. https://doi.org/10.1016/j.dib.2020.105156

Elixmann, Y., Jarke, J. (2022). Considering the Reluctance to Adopt Open Data in German Public Administration: An Exploration of Individual Innovation-Decisions. JeDEM - EJournal EDemocracy Open Gov. 14(1): 50–71. https://doi.org/10.29379/jedem.v14i1.681

Horáková, E. (2017). Open Data Potential in the Realm of Urban Planning, In PhD Research Sympozium 2017. Fakulta architektury VUT v Brne, Brno, Czech republic. 6–12. https://doi.org/10.13164/phd.fa2017.1

Kuckartz, U., Rädiker, S. (2022). Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung: Grundlagentexte Methoden, 5. Auflage. ed, Grundlagentexte Methoden. Beltz Juventa, Weinheim Basel.

Linåker, J., Runeson, P. (2020). Collaboration in Open Government Data Ecosystems: Open Cross-sector Sharing and Co-development of Data and Software. 290–303. https://doi.org/10.1007/978-3-030-57599-1_22

Máchová, R., Hub, M., Lnenicka, M. (2018). Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective. Aslib Journal of Information Management 70(3): 252–268. https://doi.org/10.1108/AJIM-02-2018-0026

Nikiforova, A. (2022). The role of open data in transforming the society to Society 5.0: a resource or a tool for SDG-compliant Smart Living? https://doi.org/10.48550/arXiv.2206.11784

Nikiforova, A., Lnenicka, M. (2021). A multi-perspective knowledge-driven approach for analysis of the demand side of the Open Government Data portal. Government Information Quarterly 38(4): 101622. https://doi.org/10.1016/j.giq.2021.101622

Norton, R.K. (2008). Using content analysis to evaluate local master plans and zoning codes. Land Use Policy 25: 432–454. https://doi.org/10.1016/j.landusepol.2007.10.006

Open Knowledge Foundation (2022). What is open? Open Data. Available at https://okfn.org/opendata/. Accessed August 30, 2022.

Publications Office of the European Union. 2022. Open data maturity report 2021. Publications Office, LU.

TRANSPARENCY IN AGRICULTURAL LAND LEASE BY LOCAL GOVERNMENT

Filip Varga[1,2], **Karlo Kević**[3], Larisa Hrustek[4*]

[1] University of Zagreb, Faculty of Agriculture, Croatia
[2] Centre of Excellence for Biodiversity and Molecular Plant Breeding (CroP-BioDiv), Croatia
[3] University of Zagreb, Faculty of Geodesy, Croatia
[4] University of Zagreb, Faculty of Organization and Informatics, Croatia

*correspondence E-mail: lhrustek@foi.hr

Keywords: open data; transparency; local government; agriculture land

1. Introduction

The demands of the information society point to the importance of public administration operating responsibly and transparently in modern digital economies. Consequently, innovations in the public sector based on the concept of 'openness' and data disclosure aim to improve transparency and accountability of power stakeholders (Park and Gil-Garcia, 2022; Lnenicka and Nikiforova, 2021). The importance of a data-driven 'transparency' approach has greatly increased greatly in the business and scientific environment (Cahlikova and Mabillard, 2020) and for the conditions, obligations and responsibilities of government in the conduct of operations and policies (Birchall, 2015). This approach contributes to good governance based on democratic values, strengthens policy making, promotes a positive impact on citizens' trust (Corrêa et al., 2017; Wiencierz and Lünich, 2022) and provides knowledge sharing for different stakeholders in the public and private sector (Park and Gil-Garcia, 2022).

Transparency implies the government's approach to providing citizens with the necessary information about decision-making, visibility and monitoring of the process as well as the outcome of government decisions (Mayernik, 2017; Lnenicka and Nikiforova, 2021; Birchall, 2015). Accountability refers to the government's ability to justify actions and decisions that affect its citizens. The operations of government institutions supported by a culture of transparency and accountability create increasing enthusiasm for the community (Cahlikova and Mabillard, 2020). Government transparency encompasses three key aspects, namely: transparency as an institutional relationship, transparency as information sharing and transparency of operations and performance (Meijer, 2013).

The implementation of government policy involves a wide range of processes and activities carried out by public institutions, and it is desirable that society is informed about them. This includes the availability of data and information on tenders, public procurement, budget management and other current issues that are important to society. Responsible and transparent management of government and public institutions is also reflected in a fair approach to the management of state

assets, e.g. state space capacity, data capacity and other resources, as well as the appropriate redistribution of all forms of state assets and resources. The creation of public value based on openness and availability of data is only possible if certain activities are transparent from start to finish.

In this research, the focus was on analysing transparency aspect of government and public administration, i.e. how transparent tenders for the allocation and disposition of state-owned agricultural land are conducted. The main objective of this work was to investigate and critically examine the practices of publishing tenders for the lease of state agricultural land in the local units of six selected counties in the Republic of Croatia.

2. Research Materials and Methods

Six counties from different parts of Croatia were selected to carry out the preliminary investigation. These are Krapina-Zagorje County, Varaždin County, Požega-Slavonia County, Lika-Senj County, Split-Dalmatia County and Istria County. The counties are distributed in different parts of the Republic of Croatia, with different levels of agricultural production and agricultural practices which ensures the relevance of the results. A total 178 local units, both municipalities and cities, were included in the analysis. The search for published tenders was conducted for the period from 2015 to September 2022.

Data collection was based on a search of the official websites of the local units. The official websites were searched in two ways: (a) checking the category "Tenders and/or Public Calls" or (b) review of the results using a search engine based on the keywords "agricultural land", "agricultural", "lease of agricultural land", "transparency". The data collected refer to tenders for the lease of state-own agricultural land. The year of publication of the tender, the link to the publication of the tender, the availability of results and decisions based on the tender and whether the local unit has a Program for the management of agricultural land was collected. In addition, data indicating irregularities in the implementation of tenders was collected, if the analysis found that tenders were repeated several times.

Data processing revealed the percentage of tenders published at the county level and the percentage of available decisions for the total number of tenders published. In addition, it was analysed how many tenders the bids of all interested parties were published transparently. Finally, the overall score of the transparency of the tenders for the lease of agricultural land was calculated. In assessing transparency and determining the overall score, the following points were taken into account: (a) whether the local unit announced the tender (1 point), (b) whether it announced the decision (1 point), and (c) whether all bids were published (1 point). In total, it was possible to score a maximum of 3 points. The total number of points achieved was used to evaluate the success rate in the responsible and transparent implementation of tenders for the lease of state agricultural land in the Republic of Croatia.

3. Results and discussion

The research results combined data for a total of 178 local units analysed, 130 municipalities and 48 cities. Of the total number of local units analysed, 54 local units were identified with published tenders for the lease of state agricultural land in the period studied. For the other 124 local units,

it cannot be said whether they have announced the tender or not, as no data is available, so further steps need to be taken to validate these results.

In the Republic of Croatia, local units are responsible for conducting tenders for the lease of state-owned agricultural land. Local units are obliged to publish the tender publicly, i.e. make it available on websites or notice boards (Ministry of Agriculture, 2019). The decisions of the local units should be made public, while there is no obligation to make the bids public.

The first part of the research involved an analysis of data at the county level. When analysing the collected data, it was found that Požega-Slavonia County had the highest percentage of published tenders for the lease of state agricultural land in the period from 2015 to 2022 (60% of local units in the county published their tenders). For the announced tenders in the local units of Požega-Slavonia County, 50% of them also published the decision on the tender. Požega-Slavonia County has the largest agricultural land capacity in the Republic of Croatia and offers space for the production of large quantities of food that could meet the needs of the population (The World Bank, 2020). Lika-Senj County was the county with the second highest percentage of published tenders (50% of the local units announced tenders, and in half of the conducted tenders the decision was also published). In Istria County, 41% of local units announced the competition and for 41% of them the final decisions were announced. Varaždin and Krapina-Zagorje counties have generally similar results, where 29% and 28% of local units respectively published tenders for the lease of state agricultural land. Varaždin County publishes the results more transparently through 38% of published decisions, while in Krapina-Zagorje County the results of only 22% of decisions on the conducted tenders are available. In Split-Dalmatia County, only 15% of local units published a tender and the results were available for 53% of the tenders conducted. Accompanying material on the details of the bids for the lease of agricultural land by interested stakeholders was available only in Požega-Slavonia County for only 33% of the conducted tenders.

The best result was achieved by Požega-Slavonia County with an average of 1.83 points, followed by Split-Dalmatia County with an average of 1.63 points and Lika-Senj County with an average of 1.50 points (**Figure 1**). It can be concluded that the mentioned counties are more successful in the transparent implementation of tenders for the lease of agricultural land. Istria County with an average of 1.41 points, Varaždin County with an average of 1.25 points and Krapina-Zagorje County with an average of 1.22 points were rated below average, i.e. it was found that the implementation of tenders for the lease of agricultural land is less transparent for the interested public. The results of this research are based on data collected from the official websites of local units, which may be one of the limitations of this research. In local units where it was not possible to collect data through this research, there are two possibilities: (a) the tender was not conducted or (b) the tender was conducted, but not published on the website. The data currently collected will be validated against the local units' responses to our inquiries about the tenders. Indeed, it is possible that some of the tenders were published on the notice boards at the location of the local units or in the official gazettes. Moreover, some of the local units have updated their official websites in recent years and there is a possibility that publications or documents have been lost in the process, i.e. the old archive is not implemented on the new official websites.

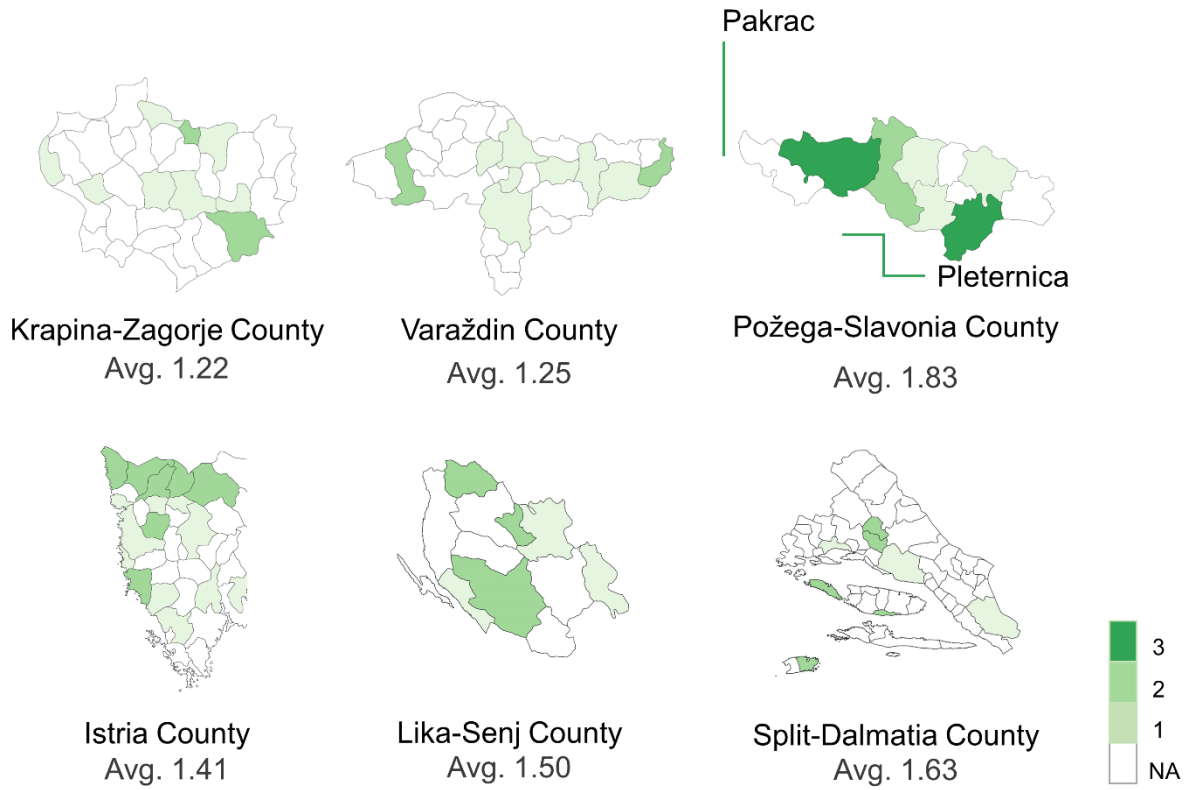

Figure 1. Transparency analysis results of public tenders for six Croatian counties.

The second part of the research involved the analysis of the collected data regarding terms of the type of local unit. It is expected that cities, as progressive local units, should be more responsible in conducting tenders by providing a transparent process. The results show that cities, as primarily urban environments, are more transparent in terms of their tenders, although there are fewer of them, i.e. 46% of the cities in the sample have announced tenders and only 25% of the municipalities. The publicly available tender results, i.e. the decisions made, followed the tenders at the level of the two types of local units to about the same extent. Thus, 41% of the cities and 44% of the municipalities announced their decision regarding their tender.

Only 9% of the tenders at the city level were followed by bids from interested stakeholders, while none were recorded for the municipalities. Although it was expected that cities would lead in the transparent conduct of public tenders, the overall score shows small differences. For cities, the overall score is 1.50 points on average, while for municipalities it is slightly lower, averaging 1.44 points.

Only two local units, the cities of Pakrac and Pleternica (Požega-Slavonia County), have published the complete documentation of their tenders, including bids from interested parties. The local units are not obliged to publicly disclose the bids received. There are concerns regarding the General Data Protection Regulation (GDPR), as the bids contain personal data of private or public persons involved in the tender. Furthermore, the publication of all documents represents an additional effort for the local units and also requires knowledge and skills in dealing with the new GDPR regulation.

Visibility of the whole tender process and availability of documentation would reduce the possibility of fraud in local units. The results show that there is evidence of fraud and manipulation in tenders in some local units, which is confirmed by several newspaper articles investigating them (Mandarić, 2022a; Mandarić, 2022b).

4 .Conclusions and future work

At the level of the Republic of Croatia, the analysis of a representative sample of local units included in the research showed an average transparency score of 1.44 points (out of 3). This means that while there is awareness of the public accessibility of tenders, a more mature level of transparency, such as publication of decisions, is still lagging behind. In addition, one-third of local units announced a tender for the lease of state-owned agricultural land, and of the total number of tenders conducted, only 43% had a decision on the selected bid. In only 4% of the tenders conducted were the bids of all interested parties presented transparently. Even though the publishing of this data is in contradiction with the GDPR regulations and local authorities are not obliged to open it, opening it is considered the highest level of transparency. The fact that cities that open this data exist, proves that it is possible to achieve the highest level of transparency and not violate personal data regulations. Based on the results, it can be concluded that there are certain obstacles that prevent the full transparency of tenders for the lease of agricultural land but it also proves local governments are familiar with the presence of open data regulations.

Future work will include the analysis of all counties in the Republic of Croatia and investigate how transparent the entire tender process for the lease of state-owned agricultural land is. The mentioned limitations in the conducted research will be removed in the next phase, as the local units have been contacted and the data that is the subject of this research has been requested from them.

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

Birchall, C. (2015). 'Data.gov-in-a-box': Delimiting transparency. European Journal of Social Theory 18(2): 185–202. https://doi.org/10.1177/1368431014555259

Cahlikova, T., and Mabillard, V. (2020). Open Data and Transparency: Opportunities and Challenges in the Swiss Context. Public Performance & Management Review 43(3): 662–686. https://doi.org/10.1080/15309576.2019.1657914

Corrêa, A. Sh., Paula, E. C. de, Corrêa, P. L. P., Silva, F. S. C. da. (2017). Transparency and open government data: A wide national assessment of data openness in Brazilian local governments. Transforming Government: People, Process and Policy 11(1): 58–78. https://doi.org/10.1108/TG-12-2015-0052

Lnenicka, M., Nikiforova, A. (2021). Transparency-by-design: What is the role of open data portals? Telematics and Informatics 61: 101605. https://doi.org/10.1016/j.tele.2021.101605

Mayernik, M. S. (2017). Open data: Accountability and transparency. Big Data & Society 4(2): 2053951717718853. https://doi.org/10.1177/2053951717718853

Meijer, A. (2013). Understanding the Complex Dynamics of Transparency. Public Administration Review 73(3): 429–439. https://doi.org/10.1111/puar.12032

Ministry of Agriculture (2019). Pravilnik o provođenju javnog natječaja za zakup poljoprivrednog zemljišta i zakup za ribnjake u vlasništvu Republike Hrvatske. Available at https://narodne-novine.nn.hr/clanci/sluzbeni/2022_09_104_1541.html. Accessed September 2, 2022.

Park, S., Gil-Garcia, J. R. (2022). Open data innovation: Visualizations and process redesign as a way to bridge the transparency-accountability gap. Government Information Quarterly 39(1): 101456. https://doi.org/10.1016/j.giq.2020.101456

Mandarić, A. (2022). Nećak Franje Lucića, Branko P. ostvario pravo na zakup više od 25 hektara zemlje u Pleternici na 25 godina?! 034 portal. Available at https://www.034portal.hr/necak-franje-lucica--branko-p--dobio-na-zakup-vise-od-25-hektara-zemlje-u-pleternici---286. Accessed September 30, 2022.

Mandarić, A. (2022). U Velikoj poništen natječaj za zakup poljoprivrednih površina u vlasništvu RH. 034 portal. Available at https://www.034portal.hr/u-velikoj-ponisten-natjecaj-za-zakup-poljoprivrednih-povrsina-u-vlasnistvu-republike-hrvatske-752. Accessed September 30, 2022.

The World Bank. (2020). Growth and Jobs in Pannonian Croatia. Available at https://www.worldbank.org/en/country/croatia/brief/growth-and-jobs-in-slavonia-baranja-and-srijem. Accessed September 30, 2022.

Wiencierz, C., Lünich, M. (2022). Trust in open data applications through transparency. New Media & Society 24(8): 1751–1770. https://doi.org/10.1177/1461444820979708

OPEN BUDGETS: OPEN DATA AS A TOOL FOR FISCAL AND TAX TRANSPARENCY

Tereza Rogić Lugarić[1], Anamarija Musa[1], **Petra Đurman[1*]**, Irena Klemenčić[1]

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

*correspondence E-mail: pdurman@pravo.hr

Keywords: Open budget data; budget transparency; budget openness; tax transparency; local units; Croatia.

Today, the principles of government transparency, openness and responsiveness in relation to public are inevitable components of good governance in general. In area of public finances they are particularly emphasized because citizens' informing and familiarization with government revenues and expenditures represents basic mechanism for holding public authorities accountable in relation to public, but is also related to the protection of human rights taxpayer protection and possibilities for public participation. Fiscal transparency has intensely been advocated by international organizations, such as International Monetary Fund (IMF), World Bank, United Nations, and other, particularly within the initiatives such as Open Government Partnership (OGP) and other anti-corruption programmes. As a principle, it includes standards such as comprehensiveness, clarity, reliability, timeliness, and relevance of public reporting on the past, present, and future state of public finances (Bronić, 2013). Areas of fiscal transparency according to IMF's Fiscal Transparency Code include fiscal reporting, fiscal forecasting and budgeting, fiscal risk analysis and management and resource revenue management.

In area of public finances, transparency occurs in two main forms – the budget (fiscal) transparency and the tax transparency. Budget transparency is related to expenditure side, while tax transparency refers to the revenue side of public finance system. The principle of budget transparency is one of the youngest budget prinicples. Informing citizens about government spending represents a mechanism of holding government accountable for its actions, which means it is an instrument for performing public oversight. It also ensures less opportunities for corruptive and other unethical behaviour and has potential to strenghten legitimacy and public trust in political institutions.

On the other hand, tax transparency primarily focuses on enabling tax authorities to collect taxes in line with the worldwide principle and hence achieving the principle of equity in taxation. Automatic exchange of information and its accompanying instruments, such as beneficial ownership registries and Country-by-Country reporting serve to avoid the unfavorable outcome of double non-taxation in cross-border affairs by limiting tax evasion and tax avoidance to the minimum level. Abolishing bank secrecy and enhancing tax transparency is expected to significantly reduce the impact of tax havens.

In this paper, our focus is limited to the budgetary dimension – its transparency, openness and availability as open data (**Figure 1**). First, an important conceptual distinction has to be pointed put. In the literature, as well as in different international benchmarks, terms of budget (fiscal) transparency and openness are often used as synonyms. Although interrelated, they refer to different processes in relation between government and the public. The principle of transparency is related to the visibility and availability of different types of government information to the public, i.e. it connotes a one-way process of communication from government to the public (Musa, 2017). The principle of openness includes the principle of transparency, but goes even further – citizens' access to information is a prerequisite for opportunity to provide a feedback information. Therefore, openness is actually related to the possibility of public to provide their opinions, suggestions and comments to the government on different aspects of its functioning. Instruments for achieving government transparency are numerous and include catalogues, brochures, databases and other types of information available on official websites of governments organizations (proactive provision of information), but access to information can also be achieved upon request of physical or legal person for obtaining an information (passive provision of information) (Musa, 2017). Instruments for achieving government openness include different instruments for involving interested public in policy making process, such as working groups for drafting regulation, public consultations, citizens' panels, questionnaires, etc. In different countries diverse instruments are applied, and they differ in the mode of involving public, the level of inclusiveness and influence provided to the public, and other features of the participatory process (Đurman et al., in press; more on participation in public administration see: Đurman, 2020). When referring to budgetary process, e-consultations and participatory budgeting are among most commonly used instruments. Finally, open budget data embraces only government information, that is, data on budgetary process available in open formats and appropriate for re-use on behalf of interested public (users). In the following section, all three aspects – budget transparency, openness and open data – will be briefly analyzed in Croatia.

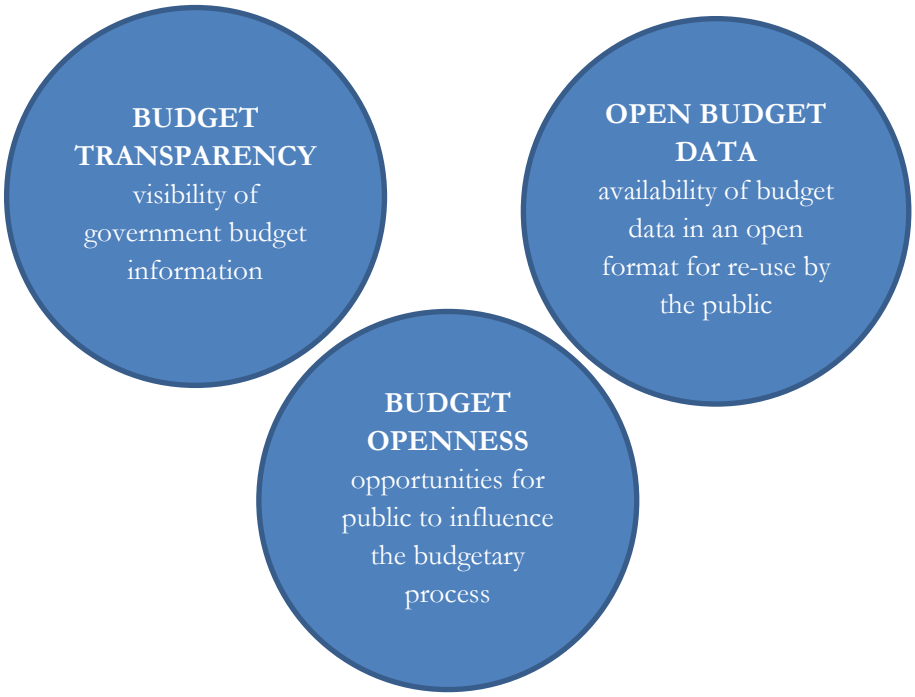

Figure 1. Budget transparency, budget openness and open budget data – conceptual distinction.

In Croatia, a legal framework for ensuring budget transparency is in place. The Law on the right to access information (LRAI) as a systemic regulation for ensuring government transparency postulates an obligation for all public authorities to publish different government acts, documents and information, while the Budget Law as a specific, sectoral law mandates transparency of financial documents. Financial acts and documents mandated to be published in open formats include budget and budgetary projections, revisions and amendments of the budget, decisions on temporary financing, reports on budget implementation, annual financial reports, strategic documents, financial plans, and public procurement procedure. According to the LRAI, financial documents and information should be published in easy searchable and machine-readable form (article 10; 28). In practice, a continuous and significant improvement in the level of government fiscal transparency has been achieved at the national, local and regional level during the last decade. This can be argued based on formal national reports, such as annual reports on the implementation of the LRAI by the Information Commissioner (specialized ombudsman for right to access information), scientific and professional analyses (such as annual analysis of the budget transparency of local and regional units in Croatia by the Institute of Public Finance), as well as some cross-national benchmarks (for example, Open budget index and others). In sum, a more consistent government transparency and openness policy is in place in general, as well as open data policy (introduced in 2018), which reflected the level of budget transparency. According to the Report of the Information Commissioner (2022), the level of proactive transparency of public authorities in general is the highest for the category of financial documents – 86.5% of all public authorities published financial reports in 2021, 85.3% published financial plans/budgets and 78% published reports on budget implementation, and there has been a continuous increase in the percentage per year in all of this categories. The level of local and regional transparency has also been increasing every year, with most transparent being counties (Croatian regional units), followed by cities (urban type of local units), while municipalities (rural type of local units) proved to be the least transparent, although with greatest disparities between them. Indicators used in the analyses by Croatian Institute of Public Finance include annual and semiannual execution of budget, budget approved and budget projections, and citizens' budget. (https://www.ijf.hr/transparentnost/?params_1=transparentnost).

With regard to openness dimension, online public consultations represent mandatory participatory instrument (according to LRAI, in all policy areas, including fiscal policy), and are largely consistently applied, especially at the national, but also at the local and regional level (with some deficiencies, though). Additional participatory instruments are facultative and rarely applied, such as participatory budgeting which has recently become more often applied instrument in cities, however, with significant differences in the implementation (see Džinić, 2021). There are some examples of 'good practice' including educational budgetary games and 'real' participatory budgeting, but they could be described more as islands of excellency than as common practice. Some international benchmark scores for Croatia point at the same direction (see **Figure 2**).

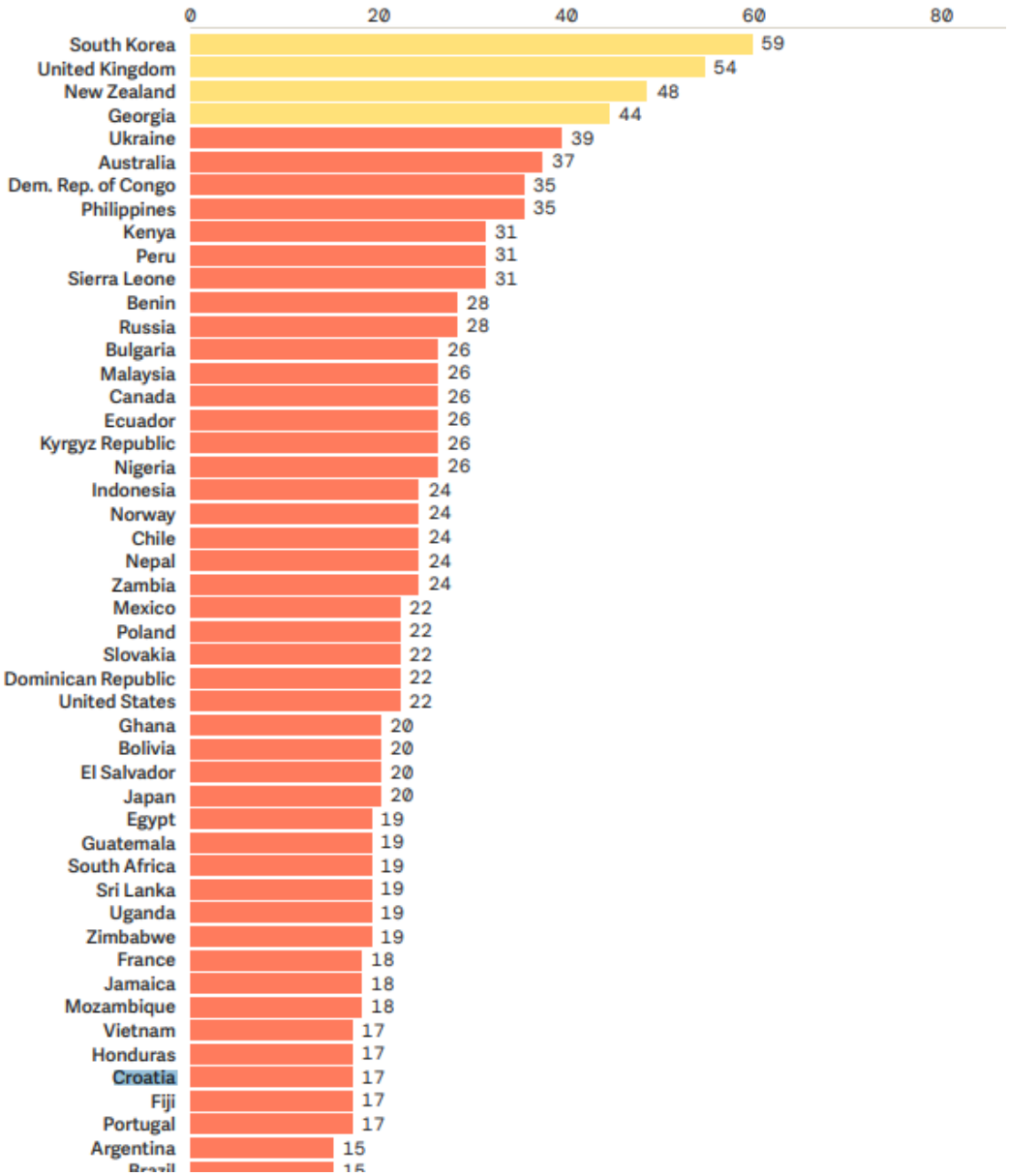

Figure 2. Croatian score on the public participation dimension according to the Open Budget Survey 2021.

With respect to open budget data in Croatia, first it has to be stressed that sectoral analyses of that open data category are lacking. Since cross-national comparisons show limited and superficial insights (predominantly focus only on transparency dimension) the aim of the paper is to investigate the availability and content characteristics of open budget data in Croatia, with special focus on local level, where greater citizens' interest for engagement can be theoretically assumed. As mentioned above, there is a legal obligation to publish budget data in an open format according to systemic law (LRAI) as well as sectoral regulation (Budget Law). Preliminary analysis of open budget data points at significant differences at national level on one hand, and local (and regional) level of government, on the other. Budgets in open formats are more often available at national level, while at the local level very diverse situation has been detected, with budget data being in some cases published in open formats (such as doc, xls, csv) as well as in other cases available in pdf or jpg formats. In addition to differences is the level of openness of budget data, there is also

an issue of content quality of budget data, which is often not formulated in a way to be clear to an average citizen/user. Recently some examples of good practice and improvements in opening budget data have been undertaken at local level, such as budget applications and efforts to achieve 'complete' budget transparency i.e. public availability of all local units' public expenditures (for example, City of Bjelovar, Split, Trogir, municipalities Omišalj, Primošten).

To conclude, main research question of the paper refers to the level of availability and quality of open budget data at local level of government in Croatia, and factors affecting the level of openness and the quality of data. In order to conduct our analysis, we rely on content analysis of budget data published by Croatian local units (cities and municipalities), with aim to analyse formats in which budget data are published, available datasets and content of budget data. Our preliminary analysis has shown that (i) legal requirements for open budget data are in place, but focused on quantitative rather than qualitative aspects of budget data; (ii) significant differences in the format of open budget data can be found between cities; and (iii) certain specificities of open data in area of public finances can be detected in relation to other categories of open data, referring to the content of open budget data and in relation to citizens' interest and familiarization with public budget.

Acknowledgements: The authors acknowledge the financial support from TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Bronić, M. (2013). Are Budgets of Croatian Local Government Units Open to Citizens?. In Musa, A. (Ed.) Transparency - why and how? 27-42. (in Croatian). Institut za javnu upravu and Friedrich Ebert Stiftung, Zagreb.

Official Gazette. (2021). Budget Law (144/21). Available at https://www.zakon.hr/z/283/Zakon-o-prora%C4%8Dunu

Đurman, P. (2020). Participation in Public Administration Revisited: Delimiting, Categorizing and Evaluating Administrative Participation. Croatian and Comparative Public Administration 20(1): 79-120. https://doi.org/10.31297/hkju.20.1.4

Đurman, P., Musa, A., Rogić Lugarić, T. (in press). The characteristics and effects of public participation in Croatian e-consultations in fiscal matters. Public Sector Journal.

Džinić, J. (2021). Participatory budgeting in the service of the development of good local governance. In Koprić, I., Staničić, F. (Eds.) Referendum and Direct Democration in Croatia 215-246 (in Croatian). Institut za javnu upravu, Zagreb.

Information Commissioner (2022). Izvješće o provedbi Zakona o pravu na pristup informacijama za 2021. godinu. Available at https://pristupinfo.hr/dokumenti-i-publikacije/izvjesca-o-provedbi-zppi/

Institute of Public Finance (2022). Proračunska transparentnost županija, gradova i općina: studeni 2021. - travanj 2022. Available at https://www.ijf.hr/transparentnost/?params_1=transparentnost

International Budget Partnership (2022). Open Budget Survey 2021. Available at https://www2.internationalbudget.org/

Official Gazette. (2022). Law on the right to access information (25/13, 85/15, 69/22). Available at https://www.zakon.hr/z/126/Zakon-o-pravu-na-pristup-informacijama

Musa, A. (2017). Information for citizens: transparent and open public administration for better governance and trust. In Koprić, I., Musa, A., Giljević, T. (Eds.) Citizens, public administration and local self- government: trust, cooperation and support 29-72 (in Croatian). Institut za javnu upravu, Zagreb.

READINESS FOR RE-USE OF OPEN GOVERNMENT DATA FOR SOLVING SOCIETAL PROBLEMS BASED ON PUBLIC PARTICIPATION: KINDERGARTENS IN THE CITY OF ZAGREB

Karlo Kević[1*], **Andrea Miletić**[1], Ana Kuveždić Divjak[1], Frederika Welle Donker[2]

[1] Chair of Geoinformation, Institute of Cartography and Photogrammetry, University of Zagreb, Faculty of Geodesy, Croatia
[2] Faculty of Architecture & the Built Environment, Delft University of Technology, Netherlands

*correspondence E-mail: karlo.kevic@geof.unizg.hr

Keywords: Open data; User perspective; Assessment; Societal problem; Kindergartens

Over the last decade, open data has become a hot topic that researchers and governments all over the world are eager to explore. The term first appeared in a 1995 document from an American scientific agency (Chignard, 2013), but its popularity increased when the U.S. president, Barrack Obama, signed the 2009 Memorandum on Transparency and Open Government, pledging to achieve an unprecedented level of openness in government (Obama, 2009). This memorandum emphasized the importance of creating a system of transparency, public participation, and collaboration through openness.

In recent years, much research has been done on the benefits arising from open government data (OGD). It is believed that OGD enables economic and social growth. According to the European Commission's 2020 report, the size of the data market in Europe is estimated to increase significantly by 2025, with possibly 2 million people employed in the open data sector. But the potential of open (government) data is going beyond just economic benefits. From a social perspective, open data enhances participation, and collaboration and empowers citizens to contribute to policies designed to their needs. Some sources state that the main motivation behind opening the data was precisely to encourage citizens to engage in solving societal problems (Huyer, 2020). This engagement is seen as a key for a successful and sustainable OGD where collaborative actions can lead to new information which could possibly solve existing societal problems (Dietrich, 2015; Purwanto et al., 2020).

For open data to reach its full potential, Pollock (2011) suggested that a traditional one-way street data model, where there is no feedback loop, and no sharing of data back to publishers and between intermediaries, should be replaced with an ecosystem approach. Charalabidis and all (2018) developed an open data life cycle model based on an open data ecosystem that brings together two main stakeholders: data providers and data users. In the inner circle are the data providers, who are responsible for creating, pre-processing, curating, storing, and publishing the data, while the outer circle consists of the data users whose role is to acquire, process, use, collaborate and provide

feedback. An important part of this process is the data re-use which is responsible for unlocking benefits arising from open data.

Different open data users can have different insights derived from open data (European Commission, 2019). These insights can accelerate the development of new applications that might provide solutions to existing societal problems. For example, in some cities, the re-use of open data on public transport accessibility ended up with new applications and services facilitating the movement of disabled people (European Commission, 2020). Since open data enables a better relationship with the government and enhances citizen participation in traditional policymaking (European Commission, 2020), this research will investigate to what extent can citizens use open data to solve a specific societal problem and provide proposals for the improvement of the city's social policy.

With the increasing urbanization and greater population migrations to the capital of Croatia, the city of Zagreb is facing a problem of discrepancy between population growth and distribution, and kindergarten infrastructure (HURO, 2022). Despite the number of kindergartens (231) (Official gazette of the City of Zagreb, 2021) in Zagreb, every year a lot of children are left unrolled or, if enrolled, they might be collocated to a random institution somewhere in the city. In the future, this problem is expected to be even more prominent with additional city population growth and revisions of existing local social policies. To mitigate this problem, the city of Zagreb made a program on preschool education for the year 2022 with a strategy on how to improve the childcare service in public kindergartens (Official gazette of the City of Zagreb, 2021). Among others, this program foresees the expansion of existing childcare capacities either by reconstruction or construction of new facilities. However, although stating where and how they plan to do it, no studies supporting decisions in the program were found. With open data initiatives supporting public participation in decision-making, this paper aims to analyze how well an end user using open data can contribute to the policy on the improvement of kindergarten infrastructure in the city of Zagreb and make it more aligned with the population's needs. To provide an answer, the analysis is divided into two parts, Data quality and availability, and Public Engagement in Policy engagement in policy-making (**Figure 1**), and focused on five specific aims:

Data quality and availability

1. Identification and analysis of sources and open government data related to population and kindergarten infrastructure
2. Analysis of existing open data assessment frameworks and detection of gaps
3. Development of a new assessment framework aligned with specific user needs
4. Analysis of open government data related to administrative boundaries, population, and kindergarten infrastructure using a newly developed assessment framework

Public Engagement in Policy-making

5. Using open data, propose improvements to the policy on kindergarten infrastructure to make it more aligned with the population's distribution and age profile.

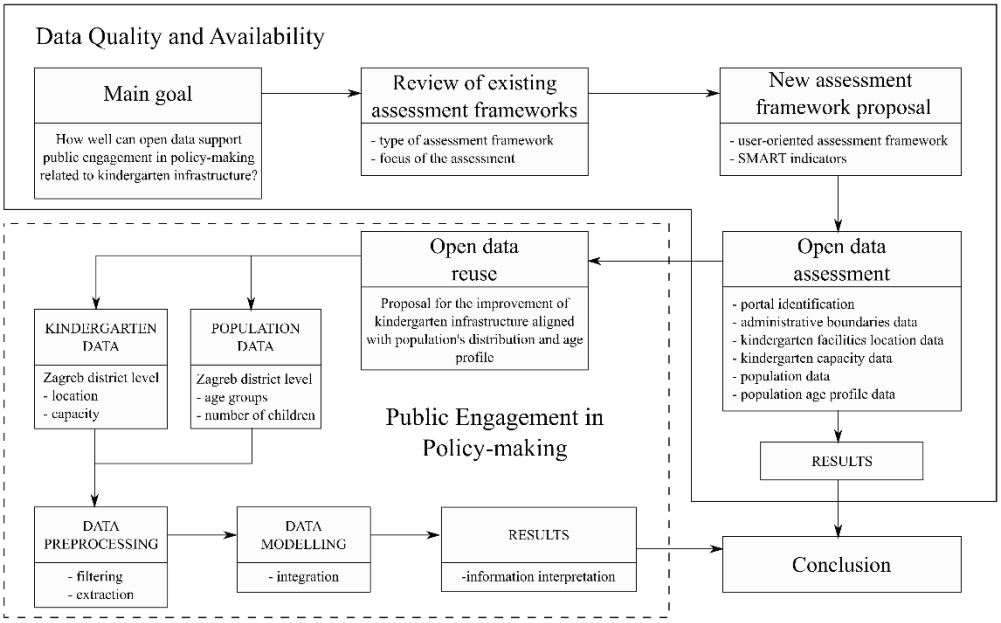

Figure 1. Data Quality and Availability Assessment

As this approach understands the re-use of open government data for solving societal problems based on public participation, the user's perspective will provide relevant and current state-of-the-art. This way, it can be assessed whether the end user, e.g., a citizen with certain data manipulation skills, can engage with data and exploit the vision of open government data initiatives. Also, providing answers to these specific aims will give information about the availability but also about suitability of open government data to solve a simple societal problem.

The methodology proposed follows the aforementioned aims. Data-quality-and-availability part of the analysis assesses the availability, findability, and level of openness of open data on administrative boundaries, kindergarten infrastructure, and statistical data about the residents of the city of Zagreb on a city district level using an assessment framework. Aligned with the aim of the analysis, the assessment framework is user-oriented meaning that its indicators are formulated to reflect on the user experience when looking for open data and based on principles of data openness. On the other hand, the public-engagement-in-policy-making part of the assessment goes deeper into the aforementioned open data and seeks if its content, resolution, and quality are appropriate to indicate the kindergarten's infrastructure alignment with the age structure of the residents in the city of Zagreb district level. This part of the analysis included data preprocessing and data modeling. Data preprocessing uses filtering to extract only data relevant to the analysis. Data modeling concerned the residents' age profile, its division into two groups: preschool children and young people, and integration with kindergarten data on the city district level. The authors consider that the division of age data into groups of preschool children and young people is relevant for two reasons. First, the number of preschool children in comparison with current kindergarten capacities will show the alignment of the service with users' needs. Second, the number of children and young people on a city district level can indicate the need for new kindergarten capacities in a specific district rather than in another district with more mature residents. Therefore, this approach is considered relevant in achieving the main goal of the analysis. In addition, with cartographic

visualization being a tool of visual communication, and having in mind its role as an intermediator between data and policymakers, results are presented on a choropleth map.

All the data needed to make this analysis are government related meaning it should be available on the National Open Data Portal which serves as a central point of access to all data related to public administration (Government of the Republic of Croatia, 2015).

This research however showed datasets to be scattered across different portals and agencies' websites, which makes it hard for users to find all the data needed and increases time spent on data searching. Data regarding the population by the city district was found on the official website of the Croatian Bureau of Statistics. However, data about the kindergartens and their location was found on the National Open Data Portal. There was also a problem with obtaining the administrative boundaries of the city districts. According to the new, European Open Data Directive (EU 1024/2019) administrative units belong to geospatial data and as such are considered high-value datasets and should be publicly available. In the case of the City of Zagreb's district administrative boundaries, no official data was found. Also, research showed that there is a limited quality of data in provided datasets; there are errors in the raw data and some data is even missing. The lack of metadata is also a big issue, for some data used, there is no information if the data is up-to-date, and even the license is not clearly determined. Based on the results it can be stated that the data is still lacking some open data aspects but however, it is considered good enough for the next phase of the analysis.

After performing the necessary steps, it was stated that the current kindergarten infrastructure does not meet public needs with the number of children outnumbering kindergartens' capacities. Also, the kindergarten infrastructure does not align with the population's age profile at the city district level. For example, the kindergarten infrastructure in the city center is not exploited enough by mostly older population identified living in the area. On the contrary, newer neighborhoods with mostly young families are lacking kindergarten space.

Seen from the results, the city of Zagreb provides open data which could be used to support public engagement in decision-making. This data is however scattered across different portals and websites and is missing metadata needed for additional context. In addition, the data is of limited quality, often with no information about the last update and terms of use. When trying to provide an answer to a simple societal question of the alignment of kindergarten infrastructure with the population's age profile, the additional problem of administrative boundaries not being officially available was faced. This suggests that some of the geospatial data, which is identified by the EU Open Data Directive as high-value datasets, is still not available for re-use. The second part of the research suggests that open data looked at is sufficient (up to a certain level) to engage the public in decision-making. However, this engagement requires end users to have certain data manipulation skills which, with all the previously mentioned limitations in data, need to be more advanced. This causes layering in defining an end user. Results of the analysis based on open data also suggest a discrepancy between kindergarten infrastructure and the population's age profile which may serve as a starting point for future policy creation by the local government.

Acknowledgments: This research is part of the "Twinning Open Data Operational – TODO" project that has received funding from the European Union's Horizon 2020 research and

innovation program under grant agreement no. 857592. Work by Andrea Miletić was additionally supported by "Young Researchers' Career Development Project – Training of Doctoral Students" (DOK-01-2020) founded by Croatian Science Foundation.

References

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., Ferro, E. (2018). The Multiple Life Cycles of Open Data Creation and Use. In: The World of Open Data. Public Administration and Information Technology, vol 28. Springer, Cham.

Chignard, S. (2013). A brief history of Open Data. ParisTech Review. Available at https://www.paristechreview.com/2013/03/29/brief-history-open-data/

City assembly of the City of Zagreb (2021). Program of public needs in preschool education and care for children of early and preschool age of the City of Zagreb for 2022 (in Croatian). Official gazette of the City of Zagreb 31/21.

Dietrich, D. (2015). The Role of Civic Tech Communities in PSI Reuse and Open Data Policies. European Public Sector Information Platform Topic Report No. 2015/05. Available at https://data.europa.eu/sites/default/files/report/2015_the_role_of_civic_tech_communities_in_psi_reuse_and_open_data_policies.pdf

European Commission (2019). Empowering people with open data. How open data enables citizens to make more informed decisions and engage with governments. data.europa.eu – The official portal for European data. Available at https://data.europa.eu/en/datastories/empowering-people-open-data

European Commission (2020). The benefits and value of open data. Building on the Digital Single Market efforts: open data in Europe today. data.europa.eu – The official portal for European data. Available at https://data.europa.eu/en/datastories/benefits-and-value-open-data

Government of the Republic of Croatia (2015). Croatian Government launches open data portal - data.gov.hr. Available at https://vlada.gov.hr/highlights-15141/archives/croatian-government-launches-open-data-portal-data-gov-hr-16574/16574

Hrvatska udruga roditelja odgojitelja – HURO (2022). Alarmantno – Gradu Zagrebu nedostaje 4030 mjesta u vrtićima! HURO. Available at https://huro.hr/alarmantno-gradu-zagrebu-nedostaje-4030-mjesta-u-vrticima/

Obama, B. (2009). Transparency and Open Government. Memorandum for the heads of executive departments and agencies. Available at https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government

Pollock, R. (2011). Building the (Open) Data Ecosystem. Open Knowledge Foundation. Available at https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/

Publications Office of the European Union, Huyer, E. (2020). The economic impact of open data: opportunities for value creation in Europe. European data portal.

Purwanto, A., Zuiderwijk, A., Janssen, M. (2020). Citizen engagement with open government data: Lessons learned from Indonesia's presidential election. Transforming Government: People, Process and Policy 14 (1): 1-30. https://doi.org/10.1108/TG-06-2019-0051

SESSION III

OPEN DATA ECOSYSTEMS

BACK TO ROOTS: RECONSIDERING DATA-AS-RESOURCE TO INTRODUCE A COMMONS-BASED MODULATING APPROACH TO OPEN DATA

Stefano Calzati[1*]

[1] Faculty of Architecture & the Built Environment, Delft University of Technology, Netherlands

*correspondence E-mail: s.calzati@tudelft.nl

Keywords: data commons; data ecosystem; data resource; open data; openness modulation, EU

This paper takes the lead from the acknowledged ongoing mismatch between open data provision and actual open data demand by users, both public and private. To tackle such mismatch, the paper suggests remodulating data openness through a commons approach, which seeks to account for all data actors' needs, while maintaining their interests in balance. In other words, data commons enacts an ecosystemic approach by default to data provision and demand, whereby data actors and their relations are mutually co-dependent. Yet, the application of the concept of the commons to data is not straightforward. This is due to data's hybrid nature, which requires reconsidering the normative understanding of data as a resource, if we really want to develop a governance model that is beneficial to all actors *and* the whole data ecosystem. Building on recent literature in this direction, the paper will briefly outline the pillars of a commons-based data ecosystem.

Since 2003 the EU (European Parliament, 2003; 2013; 2016; 2018; 2019; European Commission 2013; 2014) has released directives about the re-use of public sector information with the goal to have this information published as open data. Recently, the Data Governance Act (European Commission, 2020) and the proposed Data Act (European Commission, 2022) constitute two further pillars in this direction aimed at boosting business-to-government (B2G), citizen-to-business (C2B) and business-to-business (B2B) data sharing. While these directives signal an increasing drive towards open data and the fostering of a data-inclusive ecosystem, limitations remain. In fact, today open data are still under-exploited (Lupi et al., 2020) due also to a misalignment between the provision of data and the actual information needs of local actors. It is necessary to put equal emphasis on data supply and on data demand, for the timely release of appropriate data (Welle Donker et al., 2016). Moreover, open data initiatives have so far targeted national and supranational scales, while much data reside at local level (Verhulst, 2021). Hence, while open data represent a key institutionalized enabler to digital innovation, what is still missing are mechanisms and practices connecting (supra)nation and local levels, thus favoring the matching between provision and demand of data across scales. This is why scholars have encouraged authorities at various levels to engage citizens and promote the formation of stakeholder communities around open data (Mergel et al., 2018).

To tackle this mismatch, it is advanced here the possibility to remodulate the idea of data openness through a commons approach that can contribute to realize a sustainable open data ecosystem (van

Loenen et al., 2021). When speaking of (open) data – and more generally of data ecosystem – one should always ask: Open *how*? For *whom*? According to which *rules* and *values*? At stake is an issue not much of ownership, but control of data, latter of which is a concept open – much more than the former – to modulation. Openness, in other words, is a thick concept that lends itself and demands to be articulated on a contextual and rolling basis. As Hummels et al. (2021) note in the context of data use and reuse: "in the end, mitigation mechanisms are necessary for both those who incur damages due to their inclusion, and those who incur damages from being excluded." Concretely, this means that while the EU's striving for data openness remains the polar star, this represents an optimum to which to tend (rather than a default standard) and which requires continuous modulation to attune to and harness from all data actors' needs and provisions.

As a regime for the managing of resources, the concept of the commons can prove fruitful in this scenario. Originally, the "commons" referred to common-pool resources (CPRs) – such as fisheries or forests – characterized by non-excludability and rivalry, meaning respectively that: 1) it is difficult to forbid access and use of CPRs to any potential beneficiary; and 2) the use of CPRs depletes them and reduces further use by others. Ostrom (1990) shows that the self-management of CPRs by communities can be more efficient and sustainable than market-driven or state-led approaches, provided that formal and informal principles, roles, and rules, are designed and abided to.

By now, the commons has spilled over onto realms other than CPRs, coming to identify more broadly a system consisting of a resource, its users, the institutions binding them, and the associated mechanism processes (Feinberg et al., 2021). This characterization contains the idea that the commons exists to the extent there is a *commoning practice* (de Angelis, 2007) that conceives and manages the resource as a commons. The trading mark of such practice is to be non-appropriative by default (knowledge, technology, outputs are not owned, in the commercial sense of the term, but controlled and summoned up); collaborative by design (it considers all actors as integral to the ecosystem's flourishing), and collectively sustainable in its goals (indeed, commons for the community).

When it comes to data, data commons characterizes a regime in which actors join forces in the collection, pooling, and use of data (and infrastructures) subservient to the delivery of services for the whole community (de Lange and de Waal, 2019). As Hess and Ostrom (2007) explain, technology plays a crucial role in applying the commons to a given resource: "[t]his ability [of technology] to capture the previously uncapturable creates a fundamental change in the nature of the resource, with the resource being converted from a nonrivalrous, nonexclusive public good into a common-pool resource that needs to be managed, monitored, and protected, to ensure sustainability and preservation." This means that as soon as a (new) technology seizes or creates a resource, this can effectively be managed as a commons.

In this respect, data commons initiatives (Morozov and Bria, 2018) aim to counteract and/or repurpose the centralized ownership and use of data – either by tech companies or governments – by giving these back to people, with the goal to foster sustainable collective data practices. Data commons initiatives truly reinserts people into the data ecosystem and allow them to co-develop tech solutions with a collective socio-economic (and environmental) outlook in sight. This is why to redesign open data initiatives through a commons-driven approach might prove as a viable path to tackle the mismatch between open data provision and demand.

To do so, however, it is necessary to go back to roots and rethink our understanding of data, insofar as, while the coupling of (open) data with the commons is potentially fruitful, it is not straightforward. This is due to the current understanding of data as a resource, which is at the basis of normative approaches to data governance, including the EU's. Despite the consolidated metaphor of data as the "new oil", data are different from traditional natural resources. On this point, Prainsack (2019) notes that "although digital data clearly have material components, their materiality is of a very different kind than the physical resources." From here, it does not surprise to find both advocates of data's global nature (Shkabatur, 2019) as well as defendants of data's local rootedness (Loukissas, 2019). These positions are indeed symptomatic of the difficulty to consider data as a resource in the traditional sense of the term, that is, as a material to be found "out there" and capitalized upon. Once we dig into the nature of data-as-resource, a series of consequences derive about how to properly reconceptualize and manage data.

Differently from natural resources, data do not pre-exist in nature. Instead, data are a fully artificial (human and/or tech-created) *construct* that exists at the very moment a certain (sociotechnical) process is enacted to collect *certain* information (this is also why between information and data there is never a straightforward link, but rather a fundamental qualitative gap). This leads to suggest that data-as-resource is unique in that it manifests an *entangled nature*: if one stresses the *informational* constituency of data, then data are a virtual entity and are potentially distributable globally; if one stresses the *technical* constituency of data (from collection to storing and use), then data are material entities whose circulation can be hindered in many ways (Bates, 2018). This is why it makes little sense to ask whether data (as a resource) is scarce or abundant: data is neither scarce nor abundant – it is its use-in-context that (should) always dictate its provision. As an example, take the General Data Protection Regulation (European Parliament, 2016): a strict interpretation of the GDPR (as the "law of everything", see Purtova, 2018) signals an unbalance towards the informational side of data; a too loose interpretation of the GDPR foregrounds the technical side of data. To be sure, the entangled nature of data is also at the basis of the conflicting approaches to data ownership (Hummels et al., 2021), depending on which side is given prominence.

Therefore, back to the commons, it is somewhat misleading to ask whether data-as-resource is non/excludable and/or non/rivalrous. On the one hand, being informational, data can be accessed (in principle) by anyone, thus making them non-excludable; being technical, their access can be prevented, thus making them an excludable resource; on the other hand, being virtual, data are non-rivalrous, while as a technical artefact that contains certain information they are rivalrous. The managing of data as a commons, then, requires a paradigmatic shift in the way we think about data, passing from data as a "thing" to data as a (sociotechnical) *process*: data are always created under certain conditions, used for certain purposes, in certain contexts, and with certain results. Governance mechanisms must be designed to negotiate between the two constituencies of data and/or disentangle and give priority to either one of the two. To really enact a match between provision and demand of open data, data governance needs to move away from approaches that either target certain actors over others – e.g., citizens, public actors, private actors – or prioritize one value over others – oftentimes economic competitiveness over social inclusion or sustainability – to rather adopt an ecosystemic standpoint (Jarke et al., 2019) which, by definition, is irreducible to any of its components or relations and requires an integrated approach to find a balance of/for

the whole. In this respect, recent literature on data commons can provide a valuable starting point for advancing an ecosystemic remodulation of open data.

In her work on CPRs, Ostrom (1990) identified eight principles that allow for an effective self-management of the commons: 1) well-established boundaries of the resource; 2) rules that are context-sensitive; 3) mechanisms of collective participation for the modification of the rules; 4) these mechanisms need to be respected by external actors; 5) monitoring system for the compliance to the rules; 6) mechanisms for arbitration and sanctioning; 7) recognized and easy mechanisms for appeal; and 8) tiered management for large resources. Bloom et al. (2021) have delved into Ostrom's principles suggesting how they might be transposed in the context of data initiatives. In so doing, these authors outline guidelines concerning the governance that such commons-based data initiatives might take. If we apply to this work an ecosystemic vision, then the remodulation of open data as a commons gets inscribed into a framework whose key feature is to maintain an equilibrium among all the principles. Notably, it is possible to identify three core pillars as enablers of such equilibrium: a public-led data trust (PDT) coupled with voluntary citizen-pooled "data communes"; processes of data stewardship, and processes of data arbitration. This three-folded articulation follows up on data republicanism (Susskind, 2022), as an approach to data governance which prevents the concentration of unaccountable data power, favoring its distribution according to mutual checks and balances.

Principle 1 is the most problematic because it taps directly into the nature of the resource (i.e. data) alongside its supporting infrastructure. While data always belong to a given community/context (Bechwith et al., 2019), such community/context is entangled with both the infrastructure that supports the data and the legal framework that defines their global flow. More so: the three – data, infrastructure/legal framework, community – are always redefining poles of a whole ecosystem in the making. And it is this ecosystem in the making that needs to be "commoned". This is why the release alone of data as open data is not enough for fostering a sustainable data ecosystem; rather, rules and mechanisms are needed in the direction of a fit-for-context modulation of the control of data-as-resource (principle 2).

Principles 3 and 4 can be enacted by the constitution of a PDT which capitalizes on the institutionalization of open data, while also guaranteeing accountability and participation via a systematic opening towards diverse stakeholders. In the words of Micheli et al. (2020) a PDT "refer[s] to a model of data governance in which a public actor accesses, aggregates and uses data about its citizens, including data held by commercial entities, with which it establishes a relationship of trust." Hence, a PDT creates the conditions, under certain rules, for the commoning of data provided by a diverse array of actors. Yet, to avoid the locking up of the PDT in a form of institutional self-referentiality, data communes are also envisioned. A data commune is composed of citizens who aggregate on a voluntary basis for having their voice heard about a specific (data-related) issue. To have a data commune, the (self)identified group gathers, collects data relevant to the issue to be solved, and then asks to be formally recognized by the PDT. The recognition of the data commune, based on the provision of quality data, allows the data commune to become part of the PDT.

Since barriers to data commons initiatives often come in the forms of limited data literacy and tech capabilities (Monge et al., 2022), the constitution of data communes is supported through

institutionalized processes of data stewardship (principle 8). Moving beyond the corporate sector, Verhulst (2021) identifies data stewards as experts "identifying opportunities for productive cross-sector collaboration and responding pro-actively to external requests for functional access to data, insights or expertise." In this respect, data stewards are key enablers for linking the PDT with data communes, as well as for contributing to data literacies in citizenry and tech-legal capabilities in the public sector.

Lastly, in order to concretize principles 5, 6 and 7, it is possible to envision processes of data arbitration, in the spirit of juries responsible for deciding upon contentious issues about data use, which can happen at various scales and across various actors. Concerning both data stewardship and data arbitration, it is worth noting that since the Open Data directive (European Parliament, 2013), the EU has acknowledged the need to enforce such processes. This was reaffirmed in latest documents: about processes of data arbitration, the Data Governance Act (European Commission, 2020) states that "any natural or legal person affected by a decision of a public sector body shall have the right to an effective judicial remedy against such decision", while the proposed Data Act (European Commission, 2022) speaks of "settlement bodies" to ensure "alternative ways of resolving domestic and cross-border disputes that arise in connection with making data available." About data stewardship, the Data Governance Act specifies the need to "designate one or more competent bodies to support the public sector bodies which grant access to the re-use of the categories of data. (…) The competent body or bodies shall have adequate legal and technical capacities and expertise to be able to comply with relevant Union or national law." Moreover, member states will have to set up "structures to support public sector bodies with technical means and legal assistance." Future research will need to explore how to properly design these processes of stewardship and arbitration to make sure that they cut across scales – from the local to the European level – and contexts, i.e. facilitating cross-boundary collaboration among member states.

Altogether, PDT+data communes, processes of data stewardship and data arbitration constitute an ideal modelling (**Figure 1**) of Ostrom's 8 principles for the managing of CPRs, applied to data initiatives. While such modelling allows, in theory, for a contingent and context-based modulation of open data, thus favoring the matching between data provision and demand, the model requires testing in practice on both open data and data commons initiatives to identify barriers and enablers to its enactment.

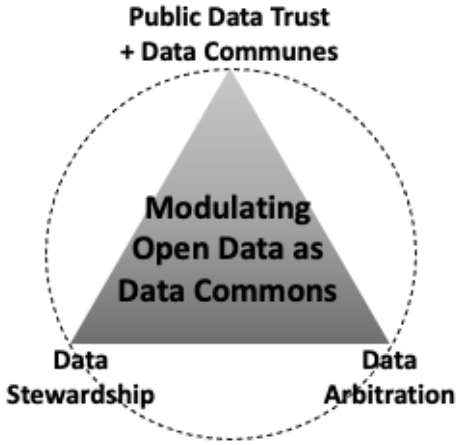

Figure 1. Modelling of open data as data commons

References

Beckwith, R., Sherry, J., & Prendergast, D. (2019). Data flow in the smart city: Open data versus the commons. In de Lange, M., de Waal, M. (Eds.) The hackable city: Digital media and collaborative city-making in the network society 205-221. Singapore: Springer.

Bloom, G., Raymond, A., Tavernier, W., Siddarth, D., Motz, G., Dulong de Rosnay, M., Ruhaak, A. (2021). A practical framework for applying Ostrom's principles to data commons governance. Available at https://foundation.mozilla.org/en/blog/a-practical-framework-for-applying-ostroms-principles-to-data-commons-governance/

de Angelis, M. (2017). Omnia sunt communia: On the commons and the transformation to postcapitalism. London: Bloomsbury Publishing.

de Lange, M., de Waal, M. (2019). The hackable city: Digital media and collaborative city-making in the network society. Singapore: Springer.

European Commission. (2022). Data Act: Proposal for a Regulation on harmonised rules on fair access to and use of data. Available at https://digital-strategy.ec.europa.eu/en/library/data-act-proposal-regulation-harmonised-rules-fair-access-and-use-data

European Commission. (2020). Data governance act. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767

European Commission. (2014). Report on high-value datasets from EU institutions. Available at http://ec.europa.eu/isa/documents/publications/report-on-high-value-datasets-from-eu-institutions_en.pdf

European Commission. (2013). Powering European public sector innovation: Towards a new architecture. Available at https://ec.europa.eu/futurium/en/system/files/ged/42-public_sector_innovation_-_towards_a_new_architecture.pdf

European Parliament (2019). Open data and the re-use of public sector information. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN

European Parliament (2018). Framework for the free flow of non-personal data in the European Union. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1807&from=EN

European Parliament (2016). General data protection regulation. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504&from=EN

European Parliament (2003). Directive on the re-use of public sector information. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32003L0098&from=en

Feinberg, A., Ghorbani, A., Herder, P. (2021). Diversity and challenges of the urban commons: A comprehensive review. International Journal of the Commons 15(1): 1-20. http://doi.org/10.5334/ijc.1033

Hess, C., Ostrom, E. (2007). Understanding Knowledge as a Commons. Baston: MIT Press.

Hummels, P., Braun, M., Dabrock, P. (2021). Own data? Ethical reflections on data ownership. Philosophy & Technology 34: 545–572. https://doi.org/10.1007/s13347-020-00404-9

Jarke, M., Otto, B., Ram, S. (2019). Data sovereignty and data space ecosystems. Business & Information Systems Engineering 61(5): 549-550. https://doi.org/10.1007/s12599-019-00614-2

Loukissas, Y. A. (2019). All data are local: Thinking critically in a data-driven society. Boston: MIT Press.

Lupi, L., Antonini, A., De Liddo, A., Motta, E. (2020). Actionable open data: Connecting city data to local actions. Journal of Community Informatics 16: 3-25. https://doi.org/10.15353/joci.v16i0.3492

Mergel, I., Kleibrink, A., Srvik, J. (2018). Open data outcomes: U.S. cities between product and process innovation. Government Information Quarterly 35(4): 622–632. https://doi.org/10.1016/j.giq.2018.09.004

Micheli, M., Ponti, M., Craglia, M., Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. Big Data & Society 7(2). https://doi.org/10.1177/2053951720948087

Monge, F., Barns, S., Kattel, R., Bria, F. A new data deal: The case of Barcelona. UCL Institute for Innovation and Public Purpose. Available at https://www.ucl.ac.uk/bartlett/public-purpose/sites/bartlett_public_purpose/files/new_data_deal_barcelona_fernando_barns_kattel_and_bria_18_feb.pdf

Morozov, E., Bria, F. (2018). Rethinking the smart city. Democratizing Urban Technology. New York: Rosa Luxemburg Foundation.

Ostrom, E. (1990). Governing the commons: The evolution of institutions for collective action. Cambridge: Cambridge University Press.

Purtova, N. (2018). The law of everything. Broad concept of personal data and future of EU data protection law. Law, Innovation and Technology 10(1): 40-81. https://doi.org/10.1080/17579961.2018.1452176

Prainsack, B. (2019). Logged out: Ownership, exclusion and public value in the digital data and information commons. Big Data & Society 6(1). https://doi.org/10.1177/2053951719829773

Shkabatur, J. (2019). The global commons of data. Stanford Technology Law Review 22: 354-411.

Susskind, J. (2022). Digital technology demands a new political philosophy. Available at https://www.noemamag.com/digital-technology-demands-a-new-political-philosophy/

van Loenen, B., Zuiderwijk, A. ., Vancauwenberghe, G., Lopez-Pellicer, F. J., Mulder, I., Alexopoulos, C., Magnussen, R., Saddiqa, M., Dulong de Rosnay, M., Crompvoets, J., Polini, A., Re, B., Casiano Flores, C. (2021). Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. JeDEM - EJournal of EDemocracy and Open Government 13(2): 1-27. https://doi.org/10.29379/jedem.v13i2.644

Verhulst, S. (2021). Reimagining data responsibility: 10 new approaches toward a culture of trust in re-using data to address critical public needs. Data & Policy 3: E6. https://doi.org/10.1017/dap.2021.4

Welle Donker, F. van Loenen, B., Bregt, A. K. (2016). Open data and beyond. ISPRS International Journal of Geo-Information 5(4): 48. https://doi.org/10.3390/ijgi5040048

NCORS OPEN DATA ECOSYSTEM: BEYOND OPEN DATA!

Warakan Supinajaroen[1], Bastiaan van Loenen[1*], Willem Korthals Altes[1]

[1] Faculty of Architecture & the Built Environment, Delft University of Technology, Netherlands

*correspondence E-mail: b.vanloenen@tudelft.nl

Keywords: Open data; NCORS; GNSS; Thailand; ecosystem

1. Introduction

The concept of open data ecosystems is on the rise (Aziz et al., 2022; van Loenen et al. 2021; Pollock, 2011). van Loenen et al. (2018) define an open data ecosystem as "a cyclical, sustainable, demand-driven environment oriented around agents that are mutually interdependent in the creation and delivery of value from open data" (cf. Boley and Chang, 2007). The successes claimed in many countries around the globe convinced the government of Thailand to implement an open data policy for its National Continuously Operating Reference Stations (NCORS), a national system that processes Global Navigation Satellite System (GNSS) signals to improve the accuracy of positional data. However, after six months, the number of users remained limited to the traditional users, and the experiment was terminated. In this paper we explore explanations for the unexpected not increasing user base of open NCORS. More specific the research presented addressed the research questions:

What are the user characteristics that facilitate the use of NCORS data in Thailand? And what are the missing elements and interactions in facilitating the use of NCORS data in Thailand's rice farming as a potential use sector?

We applied a mixed method research approach, including a literature review to investigate the key concepts for the study; a case study research approach for studying the current NCORS situation in Thailand; and interview and survey instruments to study the characteristics of a promising, but so far latent NCORS user group: Thai rice farmers.

Section 2 briefly explains the fundamental key concepts for this study. In section 3, we detail the methodology applied to investigate the use of NCORS in Thailand. Section 4 presents the results of the case study, which are further discussed in section 5. Section 6 concludes with the conclusions and recommendations for further research.

2. Key concpets

2.1. Open data

The Open Knowledge Foundation has defined open data as: "Open data and content can be freely used, modified, and shared by anyone for any purpose" (opendefinition.org).

Open data has been associated with societal benefits such as increased transparency, improved economic value creation and more efficient operations both at the provider and user side. Although these benefits may be difficult to assess (Welle Donker et al., 2016), one characteristic that has been found most often after the implementation of open data is that open data results in a significant increase in the use of the data as well as its user base (van Loenen, 2018; see also https://www.pdok.nl/rapportages for the Dutch experience).

2.2. INCORS

Satellite positioning is a technology to define positions on the earth with a precision length from millimetres to meters. Global Navigation Satellite System (GNSS) is the umbrella term for the available satellite positioning systems, for example, the Global Positioning System (GPS) of the United States, the Global Navigation Satellite System (GLONASS) of Russia, Galileo of the European Union, and BeiDou of China.

CORS is a collective term for the ground station(s) that continuously observe GNSS signals. The first CORS network was established to support GPS users accessing the National Spatial Reference System (NSRS) of the United States and correction services. Later, many CORS networks were established to support the standardisation and interoperability of spatial data and other positioning activities such as autonomous vehicles, machine control, aviation, and precision agriculture (PA).

Thailand, among many countries, has established NCORS to observe and process the Global Navigation Satellite System (GNSS) signals to improve the accuracy of positional data that serve many spatially related activities of government agencies in Thailand.

2.3. INCORS in Thailand

The development of the governmental CORS networks in Thailand started in the late 1990s. Many public agencies set up CORS networks to support their missions, such as the Department of Lands (DOL)—around 181 stations (2021), the Department of Public Works and Town & Country Planning (DPT)—15 stations, and the Royal Thai Survey Department (RTSD)—80 stations. Other public organisations also arranged their CORS networks, such as the Geo-Informatics and Space Technology Development Agency (GISTDA) and the Hydro Informatics Institute (HII).

In 2017, the government established the National CORS Data Centre (NCDC) to integrate and service CORS data from public providers. NCDC will consist of around 290 stations with an average 30-80 km distance between stations.

HII and DOL have been providing their own CORS data free of charge. In comparison, for the national CORS network controlled by RTSD, after a six-month free of charge period between 2018 and 2019, a fee based policy was reintroduced. Later, the NCDC Management Working Group meeting on July 18th 2019, approved the policy to provide CORS data services free of charge in the initial stage to promote NCDC data use and review the data access policy every six months (National CORS Data Centre, 2021).

However, the use of NCORS data in Thailand, also after introducing open data policies, has been limited only to geodetic, survey and mapping, and other scientific works in the government. Only

a few cases of the use by private sectors for construction projects were found. Significant use by new users, such as users in agriculture or industries, was not found.

3. Methodology applied to better understand the lack of increase in NCORS use

To deeper understand the situation, rice farming was chosen for investigation for three reasons. First, this sector has a significant socio-economic role in Thailand. Second, the sector faces ageing farmers and labour shortages, which requires the precise and autonomous technology provided by NCORS data. Applying NCORS technology could support the sector in tackling the issues. Third, NCORS technology has proved beneficial in farming in many other countries.

First, a survey among rice farmers was conducted to acquire user characteristics involving NCORS data use. According to agricultural experts' advice, the farm's size significantly impacts the decision to adopt the technology. Therefore, farmers were divided into Small-scale farmers (S-farmers)—representing household farmers and Large-scale farm leaders (L-farm leaders)—representing the cooperative farms. Then, to explore potential explanations for the survey results, we interviewed several experts in the field.

4. Results

The survey received the complete responses of 195 L-farm leaders and 226 S-farmers, 421 in total. Most S-farmers (58%) and L-farm leaders (70%) are in their late working years and have only primary education. Both groups showed a firm intention to use (perception of use) NCORS technology. L-farm leaders showed a higher likelihood of adopting the NCORS technology. The cost of NCORS technology appeared to be acceptable among the L-farm leaders. More than half of the L-farm leaders and 44 % of S-farmers did not know about NCORS technology before. The majority are conservative (late adopters); therefore, they need to observe the use examples before deciding to use the technology themselves. The ageing farmers and a labour shortage are the driving forces to encourage the intention to adopt NCORS technology. Some survey respondents noted the absence of NCORS technology in Thailand.

The survey also illustrated that 1) farmers demand to use NCORS technology, but NCORS technology was not available in the market, 2) the respondents had limited knowledge about NCORS, and 3) annual data cost between 1-1,000 THB (30 EURO) is acceptable for most respondents.

Interestingly, the perception of use is contrary to some points of a recent study by Sayruamyat and Nadee (2020). The study suggested that farmers have yet to perceive the usefulness of digital technology in their lifestyles. Technology was considered for the next generation. This inconsistency may involve the type of technology farmers evaluated. Digital technology alone may be perceived as less useful and necessary than a specific technology subject to this study—autonomous and precise machinery. An expert of a machinery company confirmed that benefits should be explicit to the potential users. If they see the benefits, they will find a way to use them.

Meanwhile, agricultural machine companies were informed about the new open NCORS data availability, but decided not to quickly provide the technology to the market. Machine providers expressed doubts about NCORS data availability, stability, and user readiness. They considered

that the NCORS technology was too complicated for farmers. Furthermore, they lacked personnel with NCORS expertise to support customers. The field observations also revealed knowledge gaps about NCORS's relevant technology among the machine company's specialists. To provide after-sales services, they must bridge this gap. Overall, they still need more time to incorporate NCORS technology into their business.

5. Findings

The investigation revealed that, in addition to NCORS data availability and user demand, NCORS data use necessitates facilitating conditions such as access to NCORS technology and relevant NCORS knowledge. Furthermore, in addressing the NCORS data use issue, stakeholders must understand the complexities of the NCORS data ecosystem and the time aspect of policy implementation.

5.1. Access to NCORS technology and knowledge

The availability of NCORS technology is dependent on technology providers. The study discovered a gap between the demand for and supply of NCORS technology for agriculture in Thailand. The survey indicated significant potential for NCORS use in rice farming, but the demand did not excite agriculture machine companies. Even though one company was confident in the NCORS technology for rice farming, it expressed concerns about its complexity and farmers' readiness to use it. These concerns may explain the company's lacklustre efforts.

Next to equipment availability, NCORS technology experts should also be available. The field interviews and survey highlighted the lack of experience and knowledge among the machine companies, local staff of the agricultural agencies, and farmers. The findings suggested that the readiness of human resources in all parts of the NCORS value chain must be improved as part of making technology available.

NCORS' outreach program should first target human resources in machine companies and agricultural agencies that can be change agents in delivering knowledge to farmers. It is also critical to disseminate NCORS knowledge to farmers. Delivering the technology to the farmers is another challenge due to behavioural biases (Attavanich et al., 2019). The primary target groups can be the new generation of farmers, who are the technological capable human resources in agriculture (Faysse et al., 2020).

5.2. The NCORS ecosystem

Several factors in this study are beyond an NCORS data ecosystem—they are not directly controlled by the elements in the NCORS data ecosystem. An NCORS data ecosystem depends on, involves, and is linked to other external elements: the elements and functions of other ecosystems. The internal and external engagements of the NCORS data ecosystem and surrounding ecosystems must be considered to encourage the use of NCORS data (and other data). The NCORS data ecosystem coexists with other ecosystems as part of a larger ecosystem. Defining the boundary of a data ecosystem should be discussed further in light of these complexities.

5.3. Time

Stakeholders must remember that all policies and efforts take time to achieve their objectives. The research found that time delays were overlooked in the six-month open data policy trial for the NCORS network in Thailand. It was anticipated that open data would attract users in a wide variety of sectors. However, it turned out that the number of users did not increase and was restricted to those in government agencies. The provider appeared convinced that open data did not encourage the use of NCORS and decided to discontinue open data. Such a decision disregards the time required for society or a community to adapt existing processes to the new policy situation. As seen in this case, time is needed for the machine companies to acknowledge users' demands, the knowledge spreading process, and build confidence in the technology. All these delays contributed to the diffusion of NCORS data use in Thailand.

6. Conclusions and recommendations

This research aimed to explore the policy venues to encourage the use of data from Continuously Operating Reference Stations (CORS) at a national level (NCORS). Thailand, among many countries, has established NCORS to improve the accuracy of positional data that serve many spatially related activities of government agencies.

A specific user group sector was chosen for investigation – rice farming – due to its socio-economic role in Thailand and the potential of the activity in using NCORS data. The research applied a survey to investigate the farmers' needs, practices, and attitudes toward NCORS technology. The result showed that 1) most farmers were willing to use NCORS technology, but NCORS technology was not available in the market for them to purchase, 2) Around half of the farmers knew NCORS, and 3) the annual data cost between 1-1,000 THB (30 EURO) is acceptable for most respondents.

The investigation revealed that, in addition to NCORS data availability and user demand, NCORS data use necessitates the availability of facilitating conditions such as NCORS technology and relevant NCORS knowledge. This study identified three areas for improvement in encouraging NCORS data use: the availability of NCORS technology, relevant knowledge related to NCORS data use, and the cost of NCORS data use.

The first policy triggers the NCORS data use diffusion by making NCORS technology readily available. The availability of NCORS technology will allow the farmers to use NCORS data and trigger further adoption and diffusion of NCORS data use. The second policy is to accelerate the diffusion of NCORS data use by addressing the knowledge about NCORS data availability and NCORS technology. Farmers are the primary target of the knowledge campaign. In addition, the knowledge should target the change agents and opinion leaders, who have a crucial role in delivering/fostering the knowledge to/of farmers. The staff of machine companies and agricultural departments are the change agents. Successful farmers, village sages, and academics are examples of opinion leaders or influencers. The third policy is to enlarge the potential user size through Open Data NCORS (OD-NCORS). Although the annual NCORS data cost is acceptable for most farmers, the OD-NCORS is acceptable for all potential users.

These three policies should be implemented interconnectedly. More efforts are needed to facilitate conditions that empower potential users. The critical driving forces, such as political and public support and relations with machine companies, are the strategic areas to establish. Nevertheless, the stakeholders must consider that all policies and efforts take time to exhibit the expected result.

Further research is encouraged to apply the developed concept of the NCORS data ecosystem to explore NCORS in other contexts and data types. The application may provide insights to improve the NCORS and general data ecosystem concepts. For the NCORS in Thailand, further research should look into how to better understand and improve the facilitating conditions of data use to ensure the success of open NCORS data. Finally, future research into the alignment between NCORS and CCORS for the best interest of citizens will be vital since there is a possibility of the CCORS network establishment in Thailand.

Acknowledgements: This research was funded by Her Royal Highness Princess Maha Chakri Sirindhorn of Thailand and by TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Attavanich, W., Chantarat, S., Chenphuengpawn, J., Mahasuweerachai, P., Thampanishvong, K. (2019). Farms, farmers and farming: a perspective through data and behavioral insights [Discussion paper]. Puey Ungphakorn Institute for Economic Research. Available at https://www.pier.or.th/files/dp/pier_dp_122.pdf

Aziz, A., Murillo, D. J. H., Iso, J. N., Pellicer, F. J. L. (2022). Towards a sustainable Open Data Ecosystem: First steps for optimizing findability and user feedback. Jornada de Jóvenes Investigadores del I3A 10. https://doi.org/10.26754/jjii3a.20227009

Boley, H., Chang, E. (2007). Digital ecosystems: Principles and semantics. 2007 Inaugural IEEE-IES Digital EcoSystems and Technologies Conference 398. https://doi.org/10.1109/DEST.2007.372005

Faysse, N., Aguilhon, L., Phiboon, K., Purotaganon, M. (2020). Mainly farming… but what's next? The future of irrigated farms in Thailand. Journal of Rural Studies 73: 68-76. https://doi.org/10.1016/j.jrurstud.2019.12.002

National CORS Data Centre. (2021). Introduction of establishment of National CORS Data Center (NCDC) and JICA technical cooperation. National CORS Data Centre. Available at https://ncdc.in.th/portal/apps/sites/#/ncdc-eng/datasets/829b11e51bc5469c99584acb16265e3e

Pollock, R. (2011). Welfare gains from opening up Public Sector Information in the UK. University of Cambridge: Cambridge, UK. Available at https://rufuspollock.com/papers/psi_openness_gains.pdf

Sayruamyat, S., Nadee, W. (2020). Acceptance and Readiness of Thai Farmers Toward Digital Technology. In Smart Trends in Computing and Communications 75-82. Springer, Singapore. https://doi.org/10.1007/978-981-15-0077-0_8

van Loenen, B. (2018). The Development of Open Data in The Netherlands. In van Loenen, B., Vancauwenberghe, G., Crompvoets, J. (Eds.) Open Data Exposed 215-232. T.M.C. Asser Press, the Hague. https://doi.org/10.1007/978-94-6265-261-3_11

van Loenen, B., Vancauwenberghe, G., Crompvoets, J., Dalla Corte, L. (2018). Open data exposed. In van Loenen, B., Vancauwenberghe, G., Crompvoets, J. (Eds.) Open Data Exposed 1-10. T.M.C. Asser Press, the Hague. https://doi.org/10.1007/978-94-6265-261-3

van Loenen, B., Zuiderwijk, A. ., Vancauwenberghe, G., Lopez-Pellicer, F. J. ., Mulder, I., Alexopoulos, C. ., Magnussen, R., Saddiqa, M., Dulong de Rosnay, M., Crompvoets, J., Polini, A., Re, B., Casiano Flores, C. (2021). Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. JeDEM-eJournal of eDemocracy and Open Government 13(2): 1-27. https://doi.org/10.29379/jedem.v13i2.644

Welle Donker, F., van Loenen, B., Bregt, A. (2016). Open Data and Beyond. ISPRS International Journal of Geo-Information 5(4): 48. https://doi.org/10.3390/ijgi5040048

DATA INTEROPERABILITY AND THE OPEN DATA ECOSYSTEM: ROLES AND RESEARCH AREAS

Mohsan Ali[1*], Charalampos Alexopoulos[1], Yannis Charalabidis[1]

[1] University of the Aegean, Greece

*correspondence E-mail: mohsan@aegean.gr

Keywords: open data; data interoperability; open data ecosystems; value generation

1. Introduction

Sustainability and value-creation are considered important parameters to measure the success of an open data system. Unfortunately, existing open data systems are not meeting their promises to achieve a sustainable and value-based open data system. Van Loenen et al. (2021) proposed a sustainable and value-creating open data ecosystem. According to their study, the open data ecosystem needs to be user-driven, inclusive, circular, and skill-based. In this study, we will first focus on sustainable and value-creating open data ecosystems. Secondly, we will explore how interoperability plays a role in the open data ecosystem. The interoperability of open data is an important aspect when we have open data from different fields. The interoperability frameworks and open data ecosystems will be analyzed. We will draw a conclusion that how open data interoperability can affect the open data ecosystem such as whether the interoperability is the need of user-driveness, inclusiveness, circularity, or skill-based properties. In other words, if we improve interoperability, which part of the open data ecosystem will be improved in return?

The research questions are, "What are the roles of interoperability in the sustainability and value-creation of open data ecosystems found in the literature?" and second "what are the roles of open data portals in value-creating and sustainable open data ecosystem?". To support our research questions, we will do necessary literature review to find the relatedness and required knowledge about the open data ecosystem, interoperability, and open data portals. We decided to go with the systematic literature review from 2009–2021. The systematic search will be performed with the help of different keywords such as "open data ecosystem", "open data ecosystem+interoperability", and "open data ecosystem+interoperability+portal" on different research databases such as ACM Digital Library, Science Direct, Web of Science, IEEE Xplore, and Springer. We will also perform a trend analysis based on our systematic search for how these fields will improve over the coming five years.

Traditional data is being replaced by a constantly growing dataset known as "big data." The interoperability of data is the key to exploiting the value of this data (European Commission, 2020b; Umbrich et al., 2015). The value-creating and sustainable open data ecosystem has four properties, such as user-centric, circular, inclusive, and skill-based (van Loenen et al., 2021). In this study, our objective is to analyze the role of interoperability in a value-creating and sustainable open data

ecosystem and what the research areas related to interoperability need to be addressed. A systematic literature review has been conducted to support our stance. The extensive literature review will answer the question in detail. The literature review is collected from different chapters; conference papers; books; conference proceedings; articles; and a few authenticated web resources such as policy websites and government officials. Computer Science, Business and Management, Law, Economics, Engineering, Information Systems Applications (including the Internet), Artificial Intelligence, Database Management, Computers and Society, and e-Commerce/e-business are some of the disciplines and sub-disciplines that use the term open data ecosystem.

Van Loenen et al. (2021) explained the open data ecosystem as follows: it will be value-creating and sustainable if it has four properties like user-drivenness, inclusiveness, circularity, and skill-based. In this study, the interoperability role in the value-creating, participatory, and sustainable open data ecosystem will be analyzed, such as what relationships exist between the open data ecosystem and interoperability, and what problems and improvements can be made in this regard. For example, is interoperability a part of the open data ecosystem's property of being inclusive, circular, user-driven, or based on skills? And what improvements can be made in the open data ecosystem with respect to interoperability?

2. Background

2.1. Open data ecosystems' properties

• User-drivenness: The OD ecosystem is user-driven if the data is provided based on the user's needs. And it must be balancing the demand-supply match. The feedback loop should be enabled to match the demand-supply and to increase the dataset downloads (Alexopoulos et al., 2018).

• Circularity: Existing open data systems are linear because of the unequal distribution of benefits and values among participants or stakeholders. Those who put less effort into the system are rewarded with a greater return on their investment (Gascó-Hernández et al., 2018; Sieber, 2015).

• Inclusiveness: Current open data systems are exclusive since they only account for a subset of stakeholders, including data providers (e.g., governments) and data users (e.g., businesses). Open data ecosystems should have an inclusiveness property that includes all possible players, like businesses, researchers, governments, citizens, end-users, resellers, enrichers, and facilitators (van Loenen, 2018).

• Skill-based: best effort is typically used in current OD systems. Few people have the necessary expertise to implement OD. Studies (Barbero et al., 2018; Davies et al., 2019) have shown that many organizations and governments don't have the skills to make plans that make the best use of the data they have.

2.2. Open data ecosystems' properties and Interoperability

Open data in the universe is divided into four major types: open government data; open citizen data; open research data; and open business data as described by Charalabidis et al. (2018) through the Quadruple Helix-based open data ecosystem. The diverse types of stakeholders and the data from them increase the complexity of the open data ecosystem regarding the exchange of data or

interoperability of data. Interoperability has different layers according to different research conducted in the past. According to Janssen et al. (2014a), there are four layers that can be found: technical, syntactic, semantic, and pragmatic layers. According to the European Commission's report (2017), its status that there are four interoperability layers to transfer goods, services, people, and capital between the 28 member states. Interoperability considerations should be sustainable, participatory, and value-creating (**Figure 1**). They should not be just one target or project oriented (EIF, 2017; van Loenen et al., 2021).

- Legal interoperability: The open data ecosystem is made up of different organizations and stakeholders, and legal interoperability makes sure that they can legally communicate with each other by making sure that different policies, legal frameworks, and strategies can work together.

- Organizational interoperability: The responsibilities, business processes, organizational relationships, and expectations are aligned among different organizations by using the commonly agreed goals and benefits in the organizational interoperability layers (Joinup Europe, 2022).

- Semantic interoperability: Semantic interoperability makes sure that the meaning of data and information that is transferred is always understood and retained (EIF, 2017).

- Technical interoperability: Interoperability can be seen in things like interface specifications, connectivity services, data integration services, data presentation and sharing, and secure communication protocols.

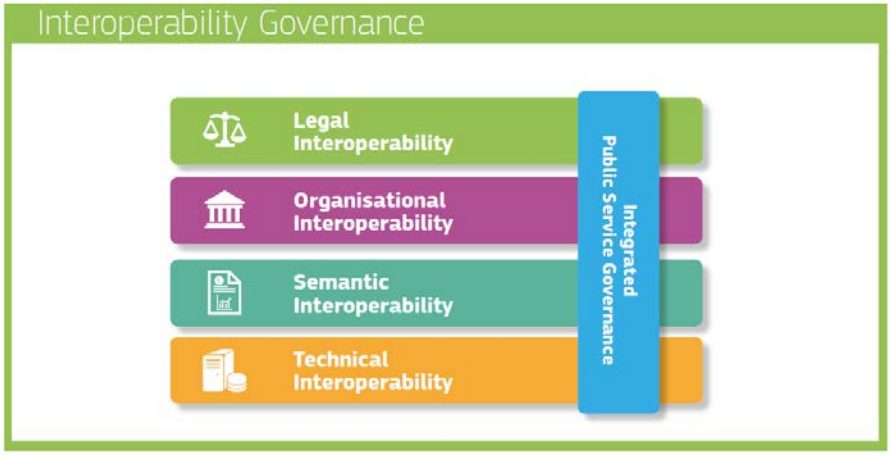

Figure 1. Interoperability model and layers (EIF, 2017).

3. Results

In **Table 1**, the results from different digital research libraries are presented. **Table 1** shows that open data interoperability has received less attention in previous works in the field of open data ecosystem. The research areas will be discussed as well in the results section. In particular, open data ecosystem + interoperability+ portals sections need special attentions.

Table 1. Research articles found in different digital libraries from 2009 to 2021

Research databases	Keyword searched	2014	2015	2016	2017	2018	2019	2020	2021	2022
Web of Science	open data ecosystem	2	1	5	1	3	1	5	4	6
	open data ecosystem + interoperability	0	0	0	0	0	0	0	1	1
	open data ecosystem + interoperability + portal	0	0	0	0	0	0	0	0	0
IEEE Xplore	open data ecosystem	2	1	3	1	0	0	0	0	0
	open data ecosystem + interoperability	0	0	0	0	0	0	0	0	0
	open data ecosystem + interoperability + portal	0	0	0	0	0	0	0	0	0
Springer	open data ecosystem	2	2	8	15	23	21	15	20	44
	open data ecosystem + interoperability	0	0	0	0	0	0	0	0	0
	open data ecosystem + interoperability + portal	0	0	0	0	0	0	0	0	0
Science Direct	open data ecosystem	0	4	6	7	1	8	4	15	4
	open data ecosystem + interoperability	0	2	2	2	2	6	3	5	4
	open data ecosystem+interoperability + portal	0	2	2	0	0	5	2	5	2
ACM Digital Library	open data ecosystem	10	1	7	6	8	0	3	7	3
	open data ecosystem + interoperability	0	0	0	0	0	0	0	0	0
	open data ecosystem+interoperability + portal	0	0	0	0	0	0	0	0	0

The results of this study have been articulated in **Table 2**, which presents the open data ecosystem properties, the role of interoperability in each property, and finally research areas identified in each property with respect to interoperability.

Table 2. Interoperability roles and research areas with respect to the properties of open data ecosystems

Open data ecosystem properties	Roles of interoperability in open data ecosystem	Research areas
User-drivenness	Organizational interoperability relates to the user-centric property of the OD ecosystem (EIF, 2017). The interoperability governance model can also be used to improve the user-drivenness of the OD ecosystems. The interoperability governance model modifications will help to devise the new interoperability frameworks, decisions, policies, roles, and responsibilities to improve national and EU level interoperability with respect to user-needs. To fix things, we need new OD governance and business models, together with technical solutions and	•Identification of needs through the data portal and usage matrix •The Open Development Interoperability governance model for the OD ecosystem •User-centric business process development for the OD ecosystem •Multi-channel data collection and data provision mechanisms •Systematic collection of users' feedback to develop the user-driven open data ecosystem •Organizational interoperability for the sustainable, value-creating, and participatory OD ecosystems •The legal interoperability of OD and the motivation behind the selection of open data licenses chosen by different stakeholders such as governments, researchers,

	tools that suggest a higher degree of government interaction with OD users.	citizens, and businesses. The legal interoperability of OD to improve the OD ecosystem's user-centric property. •A governance model for open data ecosystems and interoperability •Business models for the OD ecosystem and interoperability •Data anonymization issues and user-centric open data ecosystems (a person's identification through the OD issue)
Circularity	The definition of circularity defines the best distribution of values among the stakeholders. The involvement of stakeholders and their importance should be highlighted through the means of open data life cycle implementation as well as clearly defined roles for each stakeholder in open data portals/platforms The interoperability model needs to be circular to make the open data part circular. For instance, organizational, legal, semantic, and technical interoperability all need to be more aligned and circular among the stakeholders. At this stage, we need more research to clarify the relationship between circularity and interoperability.	• Measuring the usage of OD data from different stakeholders. For instance, the interoperability framework should record the datasets transferred to each stakeholder for better value distribution. • Transparency-by-design in the development Open data portals and interoperability of open data affect the circularity of open data. • The role of semantic interoperability in the circularity of the open data ecosystem. • technical methods to involve non-government stakeholders as data providers. • incentivizing the users based on their involvement in the open data portals. • Dataset searching algorithm • Dataset identification process based on the inner content of the data (like music search in a database).
Inclusiveness	The inclusiveness(accessibility and inclusion of stakeholders) property of the open data ecosystem is clearly explained in the principles of interoperability by the European Commission interoperability framework (EIF, 2017). Interoperability's job is to build up the whole open data ecosystem so that everyone can use and access open data everywhere in the world without any bias or discrimination.	•development of data access methods: SPARQL, bulk-downloads, big-queries development. •development of cloud-based open data portals •Use of Blockchain for OD storage to avoid the duplication of datasets •Publishing open data to multiple clouds, combining open data from multiple clouds (such as Azure and Amazon Cloud), •Better recommendation systems to improve the inclusiveness of open data stakeholders •Data integrity issues and security concerns in the interoperability of the OD ecosystem •Data distribution generations such as tracking data changes automatically •Development of on-portal data insight tools for greater inclusivity
Skill-based	Based on our best knowledge, skills are required at every stage of the interoperability framework's development. For instance, the EIF states that "the lack of the necessary in-house skill sets is another barrier to implementing interoperability policies." The skills are not limited to the implementation of policies; skills are needed to develop an interoperable open data ecosystem as well. In the open data ecosystem perspective, a lack of skill-based practices exists. That's why the existing open data systems are called best-effort-based systems (European Commission, 2020a).	•Certifications and MOOCs to prepare the stakeholders with the required interoperability skills. •Understanding of legal, organizational, semantic, and technical interoperability concepts and frameworks •Interoperability policies, barriers, and related skills •Interoperability and governance skills •Open data portal development, data provision, user-involvement, and user-interface development skills •On-portal tutorial to understand the open data portal

4. Conclusion

In this study, we analyzed the existing knowledge about the open data ecosystem and interoperability. The systematic literature review (SLR) was performed to collect the relevant research articles from 5 digital libraries using a predefined keywords list (e.g., "open data ecosystem", "open data ecosystem + interoperability", "open data ecosystem + interoperability + portal"). The role of interoperability in the sustainable, participatory, and value-creating open data ecosystem has been explained based on the literature collected from digital libraries. The explanation of the interoperability layers and OD ecosystem properties is self-evident that there is a clear gap in the interoperability layers to achieve the desired sustainable, participatory, and value-creating OD ecosystem. Furthermore, the research areas that are related to interoperability layers, principles, and recommendations in the light of the OD ecosystem properties have been recognized. The research areas are mapped to each of the properties of the open data ecosystem, which has been elaborated. Please note that the results and outcomes of this extended abstract are not completely finalized. It requires more effort to make them more precise and accurate.

Acknowledgements: This research has been financially supported by the European Union under the Marie Skłodowska-Curie ODECO project (grant agreement No. 955569) and the TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Alexopoulos, C., Loukis, E., Mouzakitis, S., Petychakis, M., Charalabidis, Y. (2018). Analysing the Characteristics of Open Government Data Sources in Greece. Journal of the Knowledge Economy 9(3): 721-753. https://doi.org/10.1007/s13132-015-0298-8

Antonio, E. C., Das, D., Bandyopadhyay, S., Feraco, A. (2017). A Practical Guide to Sentiment Analysis. https://doi.org/10.1007/978-3-319-55394-8_5

Badan Pusat Statistik. (2010). Rencana Strategis Badan Pusat Statistik.

Barbero, M., Bartz, K., Linz, F., Mauritz, S., Wauters, P., Chrzanowski, P., Osimo, D. (2018). Study to support the review of Directive 2003/98/EC on the re-use of public sector information. Available at https://ec.europa.eu/digital-single-market/en/news/impact-assessment-support-study-re-vision-public-sector-information-directive

Chandani, V., Purwanto, R. S. W. (2015). Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film. Journal of Intelligent Systems 1(1): 56–60.

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., Ferro, E. (2018). Open data interoperability. Public Administration and Information Technology 28: 75–93. https://doi.org/10.1007/978-3-319-90850-2_5

Da Silva, N. F. F., Coletta, L. F. S., Hruschka, E. R., Hruschka, E. R. (2016). Using unsupervised information to improve semi-supervised tweet sentiment classification. Information Sciences 355–356: 348–365. https://doi.org/10.1016/j.ins.2016.02.002

Davies, T., Walker, S. B., Rubinstein, M., Perini, F. (2019). The State of Open Data Histories and Horizons.

Dong-Hyun L. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. ICML 2013 Workshop: Challenges in Representation Learning, July 2013, 1–6.

EIF (2017). European Interoperability Framework. In White Pages. https://doi.org/10.2799/78681

European Commission. (2020a). eGovernment Benchmark 2020: eGovernment that works for the people.

European Commission. (2020b). Public sector information, open data and big data (RP2020).

Fudholi, D. H., Juwairi, K. P. (2009). Pemanfaatan Teknik Semi-Supervised Learning untuk Klasifikasi Dokumen Medis. Universitas Islam Indonesia.

Global Digital Insights. (2021). Digital 2021 Indonesia - We Are Social. In Global Digital Insights.

He, Y., Zhou, D. (2011). Self-training from labeled features for sentiment analysis. Information Processing and Management 47(4): 606–616. https://doi.org/10.1016/j.ipm.2010.11.003

Indrajit, A. (2018). One Data Indonesia to Support the Implementation of Open Data in Indonesia. T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-261-3

Ipmawati, J., Kusrini, Luthfi, E. T. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. Indonesian Journal on Networking and Security 6(1): 28–36. http://dx.doi.org/10.55181/ijns.v6i1.1444

Islami, M. J. (2021). Implementasi Satu Data Indonesia: Tantangan dan *Critical Success Factors* (CSFs). Jurnal Komunika: Jurnal Komunikasi, Media Dan Informatika 10(1): 13. https://doi.org/10.31504/komunika.v10i1.3750

Joinup Europe. (2022). Interoperability layers Translate. Interoperability Solutions. Available at https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/3-interoperability-layers

Ligthart, A., Catal, C., Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. Artificial Intelligence Review 54(7). Springer Netherlands. https://doi.org/10.1007/s10462-021-09973-3

Media Indonesia. (2016). Penerapan Satu Data dari BPS Mulai 2017. Media Indonesia. Available at https://mediaindonesia.com/ekonomi/42866/penerapan-satu-data-dari-bps-mulai-2017

Mila Gascó-Hernández, Erika G. Martin, Luigi Reggi, Sunyoung Pyo, L. F. L.-R. (2018). Promoting the use of open government data: Cases of training and engagement. Government Information Quarterly 35(2): 233–242. https://doi.org/10.1016/j.giq.2018.01.003

Peraturan Presiden No.39 (2019). Tahun 2019 tentang Satu Data Indonesia, Peraturan Presiden 1

Presiden RI, J. W. (2019). Pidato Kenegaraan.

Sieber, R. E., Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. Government Information Quarterly 32(3): 308–315. https://doi.org/https://doi.org/10.1016/j.giq.2015.05.003

Sekretariat Nasional Open Government Indonesia. (2020). Rencana Aksi Keterbukaan Pemerintah Indonesia 2018-2020.

Singgalen, Y. A. (2021). Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Sistematic Literature Review. Journal of Information Systems and Informatics 3(2): 278–302. https://doi.org/10.33557/journalisi.v3i2.125

Umbrich, J., Neumaier, S., Polleres, A. (2015). Quality assessment and evolution of Open Data portals. Proceedings - 2015 International Conference on Future Internet of Things and Cloud, FiCloud 2015 and 2015 International Conference on Open and Big Data, OBD 2015 404–411. https://doi.org/10.1109/FiCloud.2015.82

United Nations. (2019). United Nations National Quality Assurance Frameworks Manual for Official Statistics. In United Nations National Quality Assurance Frameworks Manual for Official Statistics. United Nations Publication. https://doi.org/10.18356/1695ffd8-en

van Loenen, B. (2018). Towards a User-Oriented Open Data Strategy. 33–53. https://doi.org/10.1007/978-94-6265-261-3_3

van Loenen, B., Zuiderwijk, A., Vancauwenberghe, G., Lopez-Pellicer, F. J., Mulder, I., Alexopoulos, C., Magnussen, R., Saddiqa, M., Dulong de Rosnay, M., Crompvoets, J., Polini, A., Re, B., Casiano Flores, C. (2021). Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. JeDEM - EJournal of EDemocracy and Open Government 13(2): 1–27. https://doi.org/10.29379/jedem.v13i2.644

Wei, Y., Gong, P., Zhang, J., Wang, L. (2021). Exploring public opinions on climate change policy in "Big Data Era"—A case study of the European Union Emission Trading System (EU-ETS) based on Twitter. Energy Policy 158: 112559. https://doi.org/10.1016/j.enpol.2021.112559

DESIGNING OPEN GOVERNMENT DATA PROGRAMS FOR USABILITY: THE IMPACT OF USABILITY ATTRIBUTES ON OPEN DATA USE

Kayla Schwoerer[1*]

[1] Department of Political Science and Public Administration, Vrije Universiteit Amsterdam, The Netherlands

*correspondence E-mail: k.n.schwoerer@vu.nl

Keywords: usability; data quality; open data use; design; Socrata Inc.

1. Introduction

Open government advocates argue for greater access to government data in an effort to increase transparency, accountability, and participation. In recent years, scholars have investigated such claims empirically, finding evidence that open government data (OGD) can indeed foster democracy (Ruijer et al., 2017; Ruijer and Martinius, 2017), provide public value (Gurin, 2014; Janssen et al., 2012), and enhance transparency and accountability (Lourenço, 2013; Lourenço et al., 2017; Lnenicka and Nikiforova, 2021; Lnenicka et al., 2021). Of course, the value that OGD provides to governments and its stakeholders is heavily contingent on its use. Yet, scholars argue that a major barrier to open data initiatives still lies in the general lack of such use (Safarov et al., 2017; Kapoor et al., 2015). This has led to a closer look at a number of factors that shape OGD use and reuse including the general functionality and usability of open data portals (Dawes et al., 2016; Machova et al., 2018; Hogan et al., 2017; Saxena, 2017; Lourenco, 2015) and the quality and usefulness of the datasets available on them (Ansari et al., 2022; Evans and Campos, 2013; Dawes et al., 2010; Martin et al., 2017; Wu et al., 2021).

For example, a handful of studies analyze factors such as completeness, availability, and timeliness (Vetro et al., 2016), arguing that such factors are necessary for open data usability. Indeed, usability is an important antecedent of open data initiatives, and, thus, the public value, broadly defined, that may be realized by these initiatives. However, much of the research to date remains largely normative. Despite the importance of questions surrounding the relationship between usability attributes and the public's use of open data, few studies have empirically examined the effect of usability attributes on user's engagement with open data. Yet, understanding how different dimensions of usability manifest on open data portals and, in turn, impact open data use, remains critical for both the evaluation of open government programs and the design of open data platforms. If open data initiatives, especially at the local level, are to achieve their potential as tools for increased transparency, accountability, collaboration, and participation, we must understand how the design and affordances of open data portals shape the usability of government information in the form of open data and, in turn, its use.

2. Research design

This mixed-methods study seeks to address this limited perspective by answering two questions. First, to what extent are governments promoting usability of government information on their open data portals? Second, how do efforts to improve usability influence open data use? To answer these questions, this research draws on the transparency and open government literature in public administration as well as emerging user experience research in the fields of marketing and psychology. A review of the literature informed a qualitative content analysis of observational data collected from 15 cities' open data portals: Albany (NY), Austin (TX), Chicago (IL), Cincinnati (OH), Kansas City (MO), Little Rock (AR), Los Angeles (CA), Nashville (TN), New Orleans (LA), New York (NY), Orlando (FL), Richmond (VA), Salt Lake City (UT), San Francisco (CA), Seattle (WA). The metadata from 5863 total published data objects (i.e. structured datasets) were analyzed to understand how dimensions of usability manifest on the cities' open data portals. The qualitative analysis identified a number of data quality and usability attributes that were then used as independent variables in a series of linear regression models. Using the number of dataset views and dataset downloads as dependent variables, the models allowed for a quantitative analysis of the relationship between governments' efforts to improve the usability of open data and actual open data use.

3. Summary of findings

Two main conclusions were drawn from the findings of this analysis. First, the results of the qualitative analysis highlighted the disparities in data quality and usability across the 15 cities. All cities used the Socrata Inc. platform which provides cities with the opportunity to leverage the design mechanisms of the Socrata interface in order to make open data more user-friendly. Despite this equal opportunity, the findings suggest that some cities invest more effort than others in presenting data in ways that are easy for users to understand, navigate, and find what they are looking for. This was evident by the noticeable differences in how cities' metadata, how much metadata they provided users with, and the quality and usability of the information. Some cities such as Austin and San Francisco consistently provided users with the information they would need to search, understand, and use the open data available. Whereas, cities like Orlando and Salt Lake City fell short when it came to the amount of information provided and the ways in which the information was presented, evidenced by the overall lack of metadata on the cities' platforms as well as the completeness of the metadata that was available. Observable differences in the quality of data weren't just evident across the cities though; there was also considerable variation in the quality and usability of the datasets that cities were publishing on their portals.

Nonetheless, the results of the quantitative analysis were consistent and clear: data quality and usability matter for open data use. The results of the analysis showed that leaving fields empty or neglecting to include keywords had significant (negative) impacts on open data use. On the other hand, relatively simple steps such as providing a dataset description or category had significant and sizable impacts on the number of views and downloads of a dataset. In some cases, a complete description increased the number of views by more than 230%. Therefore, governments' efforts to improve the usability of open data by presenting information in a clear, concise, and easy-to-understand manner can really pay off for open data use.

Still, all efforts are not equal, which was the second conclusion drawn. Findings suggest that efforts to improve intrinsic data quality such as ensuring completeness of administrative metadata, including tags and keywords, and a human-readable name and description for the data object had significant positive effects on both dataset views and downloads. Additionally, dimensions of contextual data quality that provide users with the necessary metadata to assist with discoverability and understandability of relevant data were positively associated with open data use. However, some usability attributes were significantly more effective at increasing views, a form of engagement that could be considered more passive in this context, while others were significantly more effective at increasing downloads, a more active form of engagement. The former might imply that the data are available to provide citizens with a "window" into the inner workings of the government while the latter is indicative of use and reuse that can support collaboration and innovation. Thus, cities may find that investing time and effort into improving one particular area of data quality and usability is a better use of resources than others, depending on their goals.

4. Theoretical & practical contributions of research

This research advances the literature in three main ways. First, in contrast to previous studies, this study combines observational data from the open data portals of 15 cities in the U.S. with two quantitative measures of use, views and downloads of each individual dataset, to understand how usability attributes shape open data use. To date, research relies mostly on attitudinal perceptions of public managers and citizens toward open data as well as researchers' observations of open data portals to draw conclusions about what factors are important for open data usability (Machova et al., 2018; Ojo et al., 2018). This study's research design, thus, offers the opportunity to not only explore how cities are promoting usability of open data but empirically test the relationship between the attributes of usability discussed throughout the literature and measures of actual open data use. Scholars have been calling for more empirical investigations into the many assumptions made about open data utilization (Safarov et al., 2017), including more quantitative research (Hossain et al., 2016) that aims to provide inferential analyses on open data and its outcomes (Thorsby et al., 2017). Therefore, this study aims to answer this call. Second, despite critiques that research on open data and similar ICT-based interventions lack sufficient theoretical foundation (Charalabadis et al., 2016; Begany and Gil-Garcia, 2021), using this novel research design, this study empirically tests claims made about governments' efforts to improve usability and effects on actual open data use, advancing theoretical understanding of how governments can more effectively leverage the affordances of open data programs and platforms to engage the public in ways relevant to democratic outcomes. Lastly, this study aims to inform a greater understanding of the design attributes that help or hinder user experience with open data platforms. While usability, similar to other dimensions of user experience, can be subjective and often difficult to understand without directly observing user behavior, the results of this study suggest that some best practices have considerable influence on open data use. As such, this paper concludes with a discussion of the affordances of open data portals and design principles that government agencies should keep in mind when designing, implementing, and managing open data portals. These affordances and design principles can help governments not only improve usability but also align their efforts with both broad-level and targeted strategic goals.

References

Ansari, B., Barati, M., Martin, E. G. (2022). Enhancing the usability and usefulness of open government data: A comprehensive review of the state of open government data visualization research. Government Information Quarterly 39(1): 101657. https://doi.org/10.1016/j.giq.2021.101657

Begany, G. M., Gil-Garcia, J. R. (2021). Understanding the actual use of open data: Levels of engagement and how they are related. Telematics and Informatics 63: 101673. https://doi.org/10.1016/j.tele.2021.101673

Charalabidis Y., Zuiderwijk A., Alexopoulos C., Janssen M., Lampoltshammer T., Ferro E. (2018). Open Government Data: Areas and Directions for Research. In The World of Open Data. Public Administration and Information Technology 28. Springer, Cham. https://doi-org.proxy.libraries.rutgers.edu/10.1007/978-3-319-90850-2_9

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Höchtl, J., Ferro, E. (2018). The world of open data. Public Administration and Information Technology. Cham: Springer International Publishing. 978-3.

Dawes, S. S., Vidiasova, L., Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. Government Information Quarterly 33(1): 15-27. https://doi.org/10.1016/j.giq.2016.01.003

Dawes, S. S. (2010). Stewardship and usefulness: Policy principles for information-based transparency. Government Information Quarterly 27(4): 377-383. https://doi.org/10.1016/j.giq.2010.07.001

Evans, A. M., Campos, A. (2013). Open government initiatives: Challenges of citizen participation. Journal of policy analysis and management 32: 172-185. https://doi.org/10.1002/pam.21651

Janssen, M., Charalabidis, Y., Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. Information systems management 29(4): 258-268. https://doi.org/10.1080/10580530.2012.716740

Janssen, M., Matheus, R., Longo, J., Weerakkody, V. (2017). Transparency-by-design as a foundation for open government. Transforming Government: People, Process and Policy 11(1): 2–8. https://doi.org/10.1108/TG-02-2017-0015

Lnenicka, M., Nikiforova, A. (2021). Transparency-by-design: What is the role of open data portals? Telematics and Informatics 61: 101605. https://doi.org/10.1016/j.tele.2021.101605

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. Government information quarterly 32(3): 323-332. https://doi.org/10.1016/j.giq.2015.05.006

Lourenço, R. P., Piotrowski, S., Ingrams, A. (2017). Open data driven public accountability. Transforming Government: People, Process and Policy 11(1): 42-57. https://doi.org/10.1108/TG-12-2015-0050

Máchová, R., Hub, M., Lnenicka, M. (2018). Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective. Aslib Journal of Information Management 70(3): 252-268. https://doi.org/10.1108/AJIM-02-2018-0026

Martin, E. G., Law, J., Ran, W., Helbig, N., Birkhead, G. S. (2017). Evaluating the quality and usability of open data for public health research: a systematic review of data offerings on 3 open data platforms. Journal of Public Health Management and Practice 23(4): e5-e13. https://doi.org/10.1097/PHH.0000000000000388

Nikiforova, A., Lnenicka, M. (2021). A multi-perspective knowledge-driven approach for analysis of the demand side of the Open Government Data portal. Government Information Quarterly 38(4): 101622. https://doi.org/10.1016/j.giq.2021.101622

Nikiforova, A., McBride, K. (2021). Open government data portal usability: A user-centered usability analysis of 41 open government data portals. Telematics and Informatics 58: 101539. https://doi.org/10.1016/j.tele.2020.101539

Ruijer, E., Grimmelikhuijsen, S., Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. Government Information Quarterly 34(1): 45-52. https://doi.org/10.1016/j.giq.2017.01.001

Ruijer, E. H., Martinius, E. (2017). Researching the democratic impact of open government data: A systematic literature review. Information Polity 22(4): 233-250. https://doi.org/10.3233/IP-170413

Ruijer, E., Grimmelikhuijsen, S., Van Den Berg, J., Meijer, A. (2020). Open data work: understanding open data usage from a practice lens. International Review of Administrative Sciences 86(1): 3-19. https://doi.org/10.1177/0020852317753068

Safarov, I., Meijer, A., Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. Information Polity 22(1): 1-24. https://doi.org/10.3233/IP-160012

Saxena, S., Muhammad, I. (2018). Barriers to the use of open government data in private sector and NGOs in Pakistan. Information Discovery and Delivery 46(1): 67-75. https://doi.org/10.1108/IDD-05-2017-0049

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. Government Information Quarterly 33(2): 325-337. https://doi.org/10.1016/j.giq.2016.02.001

SESSION IV

OPEN DATA APPLICATIONS

THE ANALYSIS OF AVAILABLE OPEN DATA IN THE EU FOR THE PURPOSE OF INCREASING THE SAFETY OF RAILWAY LEVEL CROSSINGS

Maja Tonec Vrančić[1*], Miroslav Vujić[1], Lucija Bukvić[1]

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

*correspondence E-mail: mtonec@fpz.unizg.hr

Keywords: mixed traffic flow; traffic open data; intelligent transport systems; traveller information systems

1. Introduction

One of the main focuses in traffic technology and research area refers to increasing the safety. In this context, the highlight is on places where traffic flows and different modes of transport mix: intersections, pedestrian crossings, railway-level crossings, etc. Although accidents at railway crossings do not occur so often, their consequences are much more severe compared to other road accidents, on personal, social, and financial level. Due to the physical "conflict" between road users and railway infrastructure, trains and train operations, railway level crossings represent one of the most complex traffic safety systems. Accidents on railways mainly occur due to the irresponsible behaviour of drivers and pedestrians, whether it is illegal movements on the railway, or non-compliance with traffic regulations and signalling at railway crossings, therefore it is necessary to devise a solution that will reliably influence their behaviour (Tey et al., 2011). The goal of this research is to analyse available data from the traffic system that can be used to create a solution that will reliably warn all road users of an approaching train. Such a solution would be based on historical data on critical points, i.e., crossings that have a history of traffic accidents (and can be classified as "black spots"), as well as real-time data on the position of road vehicles and especially trains. A solution based on historical data is applicable to all registered crossings for which there is available data, while, in addition, real-time data on train movement would increase reliability and could also be applied at unregistered crossings and in case of illegal movement on the railway.

2. Safety at rail level crossings

The safety solutions that are commonly used today at rail-level crossings refer to infrastructural interventions that change the level of the road or pedestrian crossing, or to signal warnings at rail-level crossings. There are two types of signalling: passive and active. Passive signalling includes road traffic signs (crossbucks) and crossing fences, while active signalling refers to light and sound signals and barriers that activate in the event of an oncoming train. Passive signalling is a simple and the least expensive solution, but it is far more susceptible to human disobedience than active signalling. On the other hand, even within active signalling, there is much room for improvement. The most widespread systems of active signalling is based on a sensor device placed at a certain

distance from the level crossing, which register the arrival of a train and sends information to signal-sound devices placed at the level crossing. In principle, this system is a very reliable solution, but considering the human factor, the possibility of technical failure, and the financial profitability of installing such a system at every level crossing, it is necessary to resort to simpler and more reliable solutions within the framework of Intelligent Transport Systems and widely available technologies of today.

3. Literature review – current research

Today, numerous studies are being conducted on increasing safety at railway and road crossings. Most of these research deal with the improvement of existing technical solutions for train detection, a more effective way of broadcasting warnings that leads to a reduction in speed, a more reliable transmission of information between the train and the infrastructure, warnings to the train driver about obstacles on the track, etc. Such solutions led to incremental design changes that have only marginal effects. The impact of any system on increasing traffic safety primarily depends on the impact on the road user himself, i.e., the level of his obedience to the system. Recent research shows that warning systems inside vehicles or smartphones can have a significant impact on the behaviour of drivers/pedestrians if they can be considered credible and trustworthy. A study on the impact of different ITS applications on drivers (Larue et al., 2015) was conducted using the simulation of level crossing with three types of warnings: visual, audible, and on-road valet system. The results showed that respondents are more inclined to use ITS technologies at passive crossings than at active crossings, and they prefer the system that is the easiest to use. The authors of a similar study (Landry et al., 2019) concluded, among other things, that warning systems inside vehicles have a lasting effect on driver behaviour even after the warning system is no longer presented. In a study by Salmon et al. (2016), using cognitive work analysis, the authors concluded that the level of safety at level crossings cannot be influenced only through changes at the level crossing itself, but the introduction of new ways of warning drivers and new data collection systems is necessary, or at least the integration of existing systems of different stakeholders. Through their research (Ryder et al., 2017) developed and tested (in real environment) a comprehensive in-vehicle decision support system which provides accident hotspot warnings based on location analytics applied to a national historical accident dataset. They demonstrated that in-vehicle warnings of historically dangerous locations have a significant improvement on driver behaviour over time while crossing these hotspots, they also raise awareness of the gradually reduced obedience of local drivers.

4. Relevant traffic data

Traffic is a very complex system composed of many interdependent elements. Most of these elements, or subsystems, use a particular form of information system and collection of relevant data. Today, many solutions in the field of transport depend on reliable and consistent spatial data. This research aims to detect data from the traffic system that can be used to create a solution for reliable warning to drivers about approaching train at a level crossing. For a quality result of such a system, it is necessary to connect data available from multiple sources or, in general, to connect certain safety features of the level crossing with their spatial component. Thus, the first step requires data on the locations of all registered level crossings. Furthermore, to classify crossings according to the danger criteria, data on the technical equipment of crossings and detailed historical data on accidents for each crossing are required, with an emphasis on the severity of the

consequences of the accident. In the last step, for the detection of an oncoming train, real-time data on the location of the train is needed. To connect these data in a quality way, data must be available in an open and machine-readable form. In the field of transport, there are data available at national and at the EU level, but they are quite limited. For example, according to the Official portal for European Data, of the total number of available data sets, the field of transport occupies only 3.75% (including all modes of transport), and their quality and quantity differ depending on the source. This is primarily because traffic, as a complex system, and especially the railways which are most often managed by the state, uses the collected data mostly for their own purposes and within a closed system, therefore the data is closed and owned by them, and part of the published data is in a form and quantity adapted to their needs, that is, of each subsystem separately. Thus, data sources differ in terms of functionality, characteristics, and service quality. Furthermore, the published data sets are not standardized, they are published in different forms and different languages depending on the source, and often there is no published description of the data.

5. Conclusion

The solution proposed in this paper must be seen as an upgrade of existing systems, and acceptance by drivers/pedestrians is necessary for a successful system. To influence the driver's/pedestrian behaviour, the data presented must be reliable. Furthermore, all railway level crossings need to be classified according to the level of safety, then subsequently according to the history of accidents for each one separately, that is, they should be assigned to the appropriate category and weighting factor according to the severity of the consequences, from mild to severe, respectively, based on the above, the crossing should be defined as a place of traffic flows interference with high or low risk. To achieve this goal, it is necessary to combine data sets on the location of the railway-level crossing, the technical equipment level, and the number and severity of traffic accidents at each individual railway-level crossing. To combine all this data, a cooperation, and data sharing between stakeholders from different subsystems is necessary. Also, it is necessary to point out the need to standardize such data sets and to open their availability to all interested groups (government, scientific community, industry, and the public), because existing data sets are incomplete, missing, not in formats available for download, or generally, not of the sufficient quality and quantity.

Acknowledgements: The authors acknowledge the financial support from TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Landry, S., Jeon, M., Lautala, P., Nelson, D. (2019). Design an assessment of in-vehicle auditory alerts for highway-rail grade crossings. Transportation Research Part F: Traffic Psychology and Behaviour 62: 228-245. https://doi.org/10.1016/j.trf.2018.12.024

Larue, G., Rakotonirainy, A., Haworth, N., Darvell, M. (2015). Assessing driver acceptance of Intelligent Transport Systems in the context of railway level crossings. Transportation Research Part F: Traffic Psychology and Behaviour 30: 1-13. https://doi.org/10.1016/j.trf.2015.02.003

Ryder, B., Gahr, B., Egolf, P., Dahlinger, A., Wortmann, F. (2017). Preventing Traffic Accidents with In-Vehicle Decision Support Systems - The Impact of Accident Hotspot Warnings on Driver Behaviour. Decision Support Systems 99: 64-74. https://doi.org/10.1016/j.dss.2017.05.004

Salmon, P. M., Lenne, M. G., Read, G. J., Mulvihill, C. M., Cornelissen, M., Walker, G. H., Stevens, N. (2016). More that meets the eye: using cognitive work analysis to identify design requirements for safer rail level crossing systems. Applied Ergonomics 53: 312-322. https://doi.org/10.1016/j.apergo.2015.06.021

Tey, L.-S., Ferreira, L., Wallace, A. (2011). Measuring driver responses at railway level crossings. Accident; Analysis and Prevention 43(6): 2134-2141. https://doi.org/10.1016/j.aap.2011.06.003

IDENTIFICATION OF FEATURES FOR TRAJECTORY SEGMENTATION ACCORDING TO THE TRANSPORT MODE

Martina Erdelić[1*], Tonči Carić[1], Tomislav Erdelić[1] and Nikola Mardešić[1]

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

*correspondence E-mail: merdelic@fpz.unizg.hr

Keywords: urban mobility; transport mode; smartphone sensor data; trajectory segmentation; feature selection

1. Introduction

A transport network is a complex system whose analysis requires data from transport network users. The influence of the growing degree of urbanization and population places transport as one of the main factors affecting the quality of life in large cities. Therefore, research fields such as urban mobility, energy consumption, pollution reduction and security are often the subject of scientific research. All these research fields are directly or indirectly related to the mobility of transport network users. Observing the model of human mobility through data collected in different transport environments can significantly contribute to solving transport problems.

An in-depth trajectory analysis, which can be supported by clustering or classification methods, can describe the mobility of users through the transport network. Recognition of everyday human activities during various locomotion and uses of transport modes has several applications, such as educating users to change their behavior pattern or promoting a healthier lifestyle. In addition, applications, which track user mobility, can help with public urban transport planning, smart parking, and vehicle traffic monitoring (Kalašová et al., 2021).

User trajectories often contain several connected transport modes that a user uses, moving from the origin to the destination. Hence, such trajectories need to be segmented before further processing. Therefore, there is a need to develop a trajectory segmentation method that will recognize the Mode Transfer Point (MTP), i.e., the point when one transport mode ends and starts the use of the following transport mode. Furthermore, trajectory segmentation is important for transportation network analysis because a mobility pattern is not the same for the entire trajectory but could be for some of its segments.

The accelerometer sensor is the most widely used among the sensor modalities for transport mode classification and trajectory segmentation (Hemminki et al., 2013). Global Positioning System (GPS) data-based approaches use features derived from speed, GPS coordinates, and traveled distance (Gong et al., 2018). Geographic Information System (GIS) data, such as bus or train stops, are also often included in datasets to avoid the shortage of relevant features. The main advantage of accelerometer sensor data is low power consumption, which enables continuous monitoring of

human behavior with higher sample frequency and does not depend on external signal sources. However, to avoid the shortage of GPS and GIS data, it is necessary to identify the relevant features that will better describe the user trajectory.

In this paper, several features, and their influence on the accuracy of a MTP detection were tested. The time and frequency domain of trajectory data are used for feature extraction. Features are selected using ANOVA F value and different feature sets are tested using Transition State Matrices (TSM) approach described in the research (Erdelić et al., 2022).

2. Materials and methods

In this paper, the benchmark dataset Sussex-Huawei Locomotion-Transportation (SHL) was used (Wang et al., 2019). SHL is a large-scale dataset of smartphone sensor data recorded through seven months in 2017, usually used for multimodal transportation analytics (Gjoreski et al., 2018). The dataset contains 753 h of multimodal data: car (88 h), bus (107 h), train (115 h), subway (89 h), walk (127 h), run (21 h), bike (79 h), and still (127 h). For this research, the original classes were transformed into two new classes: change of transport mode or no change of transport mode. While collecting data, each user had a device in multiple places on the body to track the movement. The mobile device orientation is not necessarily fixed. The dataset contains seven sensors, but only data collected from a 3D accelerometer, gyroscope, magnetometer, and gravity were used in this research. Data are collected from all sensors with a frequency of 100 Hz. The dataset is divided into training and testing datasets containing 70% and 30% of the data, respectively.

Data quality affects the data mining results; therefore, data need to be preprocessed before feature extraction. The paper presents three primary data preprocessing techniques: data cleaning, time window segmentation and data transformation using Discrete Fourier Transform (DFT). Before implementing a methodology for MTP detection, the essential part is extracting valuable and distinct information from time windows. Therefore, data are transformed from the time to the frequency domain for each sensor. Furthermore, a challenging task in the raw signal data analysis is to solve the orientation issue. In the SHL dataset, smartphones can be carried in different orientations, meaning that the same sensor values in various circumstances can represent moving in different directions. In this paper, we computed the magnitude value for all sensors by Equation 1 where $\boldsymbol{M}$ is the magnitude, and $\boldsymbol{S_x}$, $\boldsymbol{S_y}$, and $\boldsymbol{S_z}$ are sensor values along the x, y, and z axes, respectively.

$$\boldsymbol{M = S_x + S_y + S_z} \quad (1)$$

In summary, for every time window, features are computed as a compound of the three axes (x, y, and z) of sensor data with a magnitude vector. For each sensor, all features presented in Table 1 are computed. The calculated features refer to basic statistical features such as mean, median, standard deviation, variance, mean absolute value, interquartile range, maximum and minimum values, and root mean square value. Measures such as the kurtosis, skewness and peak value describe data distribution within a time window. The z value shows the distance of the data in the current time window from the mean, where the distance is measured in standard deviations. The kurtosis, skewness, and shape factor provide insight into the behavior of the data in a single time

window. In contrast, the impulse, crest, and clearness factors focus on peak values within a time window (**Table 1**).

Table 1. Computed features in the time and frequency domain

Feature	Symbol	Time domain	Frequency domain
Mean	$\overline{x}$	✓	✓
Variance	σ	✓	✓
Standard deviation	s	✓	✓
Median	X	✓	✓
Skewness	g	✓	
Kurtosis	k	✓	
Z score	z_i	✓	
Interquartile range	IQR	✓	
Peak to peak	PTP	✓	
Maximum	x_{max}	✓	
Minimum	x_{min}	✓	
Energy	E	✓	
Average absolute value	$\overline{x}_{abs}$	✓	
Root mean square	RMS	✓	
Variance of frequency data	RVS		✓
Frequency center	F_c		✓
Impulse Factor	F_i	✓	
Crest Factor	F_p	✓	
Clearance Factor	F_r	✓	
Shape Factor	F_s	✓	✓
Kurtosis Factor	F_k	✓	
Skewness Factor	F_g	✓	

All features are calculated for accelerometer, gyroscope, gravity, and magnetometer data resulting in 432 features (27 features * (3 axes + 1 magnitude) * 4 sensors= 432). To identify relevant features, feature importance analysis is performed. Another reason for feature analysis is to determine which features are more important for MTP identification, primarily due to the imbalance between the number of time windows in which MTP has occurred and those representing time windows in which there is no MTP.

The process of identifying relevant features is divided into two parts. First, the features are sorted according to the ANOVA F value, and secondly, features with a significant impact on the detection of false positives are removed (classification of time windows in which no change occurred as those in which a change occurred).

The ANOVA test is a statistical method used to determine significant differences between two or more classes by testing differences in mean values using the variance of included features (Sawilowsky, 2002). In other words, the value obtained by the ANOVA test shows how well a feature distinguishes two classes. Two measures are used to calculate the ANOVA test: variability within and between classes. Both measures use the sum of squares to determine the dispersion in the data. Variability within classes SS_u shows the dispersion of data in the entire dataset, where data belonging to different classes are considered separately, given by Equation 2. Variability

between classes SS_i shows the variability of a subset of data belonging to the same class in relation to the entire dataset, given by Equation 3. In both equations, Equation 2 and Equation 3, k represents the number of classes, x_{ij} is the i value of the observed feature belonging to the class j in a dataset of size N. $\overline{x}$ is the mean value of the observed feature, and $\overline{x_j}$ is the mean value of all records belonging to class j. The total number of records of a class is denoted by n_j. By setting the obtained values for SS_u and SS_i in a ratio, the ANOVA F value is obtained, given by Equation 4, (Sawilowsky, 2002).

$$SS_u = \frac{\sum_{i=1}^{N}\sum_{j=1}^{k}(x_{ij}-\overline{x_j})^2}{N-k} \quad (2)$$

$$SS_i = \frac{\sum_{j=1}^{k} n_j(\overline{x_j}-\overline{x})^2}{k-1} \quad (3)$$

$$F = \frac{SS_u}{SS_i} \quad (4)$$

3. Results and discussion

The dataset contains 23544 records and 547 time windows with MTP. The influence of individual feature and sensor on the trajectory segmentation result was analyzed to find features that better describe the original data. ANOVA F values are assigned to the features indicating their relevance, and the features that provide the most information during the transition between transport modes are highlighted. **Figure 4** shows feature importance, represented as filled circles with two attributes: size, which represents the importance of the feature; the greater the importance is, the larger the circle is and color, from dark blue to dark red, represents the ANOVA F value. The y-axis represents sensors, and the x-axis represents feature symbols. The best ANOVA F score is achieved for the standard deviation and the gravity along the z-axis. Of all sensors, the magnetometer shows the lowest ANOVA F values, while features with a minor influence are the median, skewness, and kurtosis.

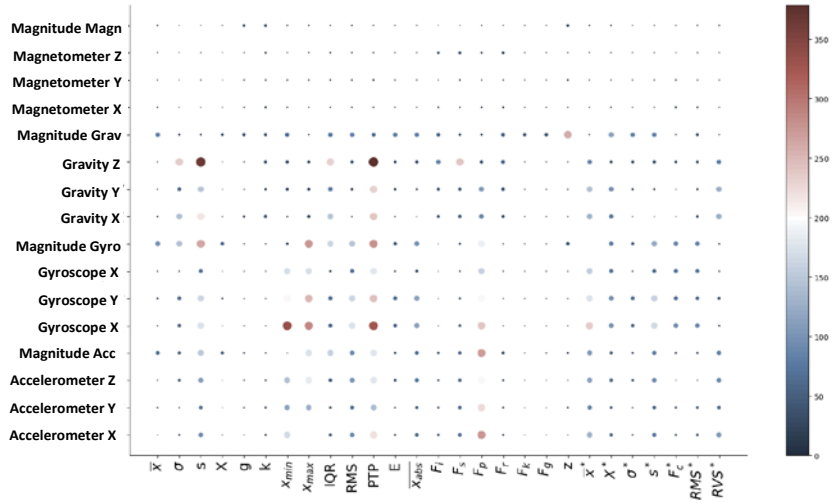

Figure 4. Feature importance using ANOVA F score

In the next step, the required number of features that will be used to predict MTP is determined using the TSM method. The testing process included 200 features with the highest ANOVA F value. Prediction accuracy was observed through measures of the model's overall accuracy, recall

and F score. **Figure 5** shows the testing result for different number of features. The best result is achieved with 80 features (the column marked with a bold black line) where the recall takes one of the highest values, and the decrease in overall accuracy is the smallest compared to the shown examples.

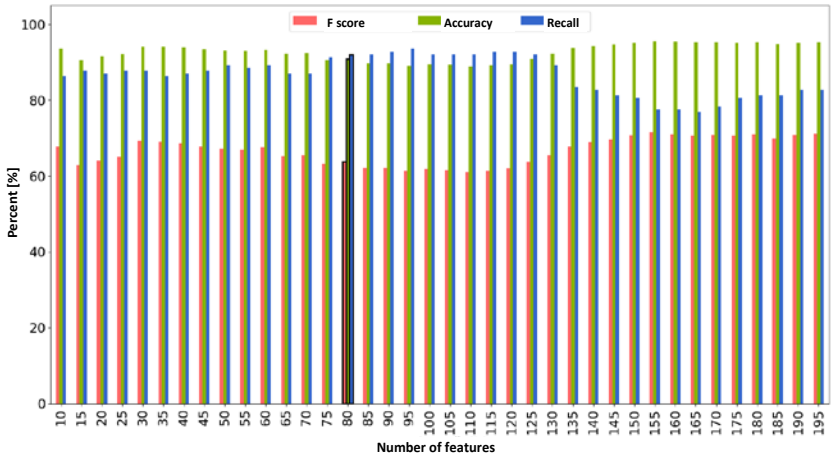

Figure 5. TSM test result for different number of features

Based on the results of the trajectory segmentation method using the identified set of features, it can be concluded that the computed and selected features can describe the user behavior pattern when changing the transport mode. However, to verify the pattern of user behavior and the applicability on other traffic networks, it is necessary to expand the dataset using open datasets on which the validation process can be performed.

Acknowledgements: The authors acknowledge the financial support from TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Erdelić, M., Carić, T., Erdelić, T., & Tišljarić, L. (2022). Transition State Matrices Approach for Trajectory Segmentation Based on Transport Mode Change Criteria. Sustainability 14: 2756. https://doi.org/10.3390/su14052756

Gjoreski, H., Ciliberto, M., Wang, L., Morales, O., Javier, F., Mekki, S., Roggen, D. (2018). The University of Sussex-Huawei Locomotion and Transportation Dataset for Multimodal Analytics With Mobile Devices. IEEE 42592-42604.

Gong, L., Kanamori, R., Yamamoto, T. (2018). Data selection in machine learning for identifying trip purposes and travel modes from longitudinal GPS data collection lasting for seasons. Travel Behaviour and Society 11: 131-140. https://doi.org/10.1016/j.tbs.2017.03.004

Hemminki, S., Nurmi, P., Tarkoma, S. (2013). Accelerometer-based transportation mode detection on smartphones. Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems. Roma, Italy.

Kalašová, A., Čulík, K., Poliak, M., Otahálová, Z. (2021). Smart Parking Applications and Its Efficiency. Sustainability 13: 6031. https://doi.org/10.3390/su13116031

Sawilowsky, S. S. (2002). Fermat, Schubert, Einstein, and Behrens-Fisher: the probable difference between two means when s1 2= s2 2. Journal of Modern Applied Statistical Methods 1(2): 461-472.

Wang, L., Gjoreski, H., Ciliberto, M., Mekki, S., Valentin, S., Roggen, D. (2019). Enabling Reproducible Research in Sensor-Based Transportation Mode Recognition With the Sussex-Huawei Dataset. IEEE 10870-10891.

OPEN SCIENCE IN MARINE ROBOTICS

Simona Aracri[1*], Roberta Ferretti[1], Corrado Motta[1], **Fausto Ferreira[2]**, Marco Bibuli[1], Francesca de Pascalis[3], Angelo Odetti[1], Gabriele Bruzzone[1], Massimo Caccia[1]

[1] CNR-INM, Uos Genoa, Italy
[2] Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia
[3] National Research Council Institute of Marine Sciences, Italy

*correspondence E-mail: simona.aracri@cnr.it

Key words: Marine Robotics; Open Science; FAIR data; Environmental Monitoring; Emerging Technology

1. Introduction

The concept of Open Science translates in sharing knowledge for progress and for the general public benefit. In essence, doing Open Science means sharing every step of the research process, from data, to protocols, to software, to results. The ultimate goal remains to share knowledge as quickly, broadly and effectively as possible.

The benefits of Open Science are multiple; for science itself, which becomes more transparent, verifiable, reproducible as well as faster and efficient, and therefore contributes to an acceleration in the process of creating the knowledge; for companies, which can take advantage of research results and, combining them with their specific expertise, offer more innovative products; for the whole society, in particular citizens and administrators, who can make decisions more objectively based on data (Fritz et al., 2019) and their scientific interpretation.

Since data are essential to the scientific process, Open Science efforts have focused on making data more openly available. Open data are data that may be accessed, used, and shared for any purpose without restrictions. However, making open data available is not enough. It has to be done in an intelligent way, for example making data available in an open repository and in a nonproprietary, standardized format. Whenever possible, a DOI should be created for the data together with a clear licensing information along with use constraints about the data (Ramachandran et al., 2021). All these good practices are part of the more general FAIR guiding principles for scientific data management (Wilkinson et al., 2016). Following FAIR principles, gathered data can be used in multiple fields of science not necessarily only within the collecting community, which is more likely to share the same vocabulary and background knowledge. Without FAIR research data, open science is simply impossible (Burgelman et al., 2019).

1.1. The key role of metadata

Research data needs metadata to become Findable, Accessible, Interoperable and Reusable (FAIR), by humans and machines (Wilkinson et al., 2016). The first step to render data FAIR is to establish

a set of descriptive metadata, i.e. information to make the dataset findable by the wider scientific community. Usually includes information such as title, author, subjects, keywords, publisher, urls and are typically domain agnostic. Different standards are already consolidated for the descriptive metadata such as ISO19115 (ISO19115 standard), Dublin core (DCMI, 2022), and DataCite (DataCite Schema) among others.

In marine robotics, while there are a few efforts in standardization (Waldmann et al., 2021) mostly these are either for industrial applications such as Remotely Operated Vehicles (ROVs) (NORSOK Standard), military-originated (NATO, STANDARD ANEP-87) or simply best practice community efforts (Hörstmann et al., 2020). Moreover, these standardization efforts are related to the construction or operation of robots or their scientific instrumentation. On the other hand, much less attention has been paid to the need for metadata standards and their importance as a cornerstone for open data. While there are a few efforts such as Marine Regions (https://www.marineregions.org) for what concerns georeferencing of marine areas (e.g. for trials) or the vocabularies defined in the NERC Vocabulary Server (The NERC Vocabulary Server (NVS)), most marine robotics research groups are not using shared metadata standards (sometimes not even ISO 8601 (ISO, 2022) for Date & Time). This is detrimental to fruitful collaborations among different research groups as well as to the re-use of published open data. The Robotic Operating System (ROS) has become a standard middleware among the marine robotics community, but this does not include any metadata standardization and different groups attribute different terminologies to the same data (and metadata). Thus, a collaborative effort is needed to define shared and easy adoptable metadata standards and a shared vocabulary in the field of marine robotics. Without these, open data will suffer from lack of findability and interoperability violating two of the FAIR principles. This need is exacerbated by the open data mandates by several funding institutions such as the European Commission or UK Research and Innovation. Therefore, rather sooner than later the marine robotics community needs to address the lack of common metadata, shared vocabulary and data format to open their data to the public. One of the few examples available of a marine robotics open dataset properly documented in a data paper can be found in (Bernardi et al., 2022).

Working to set the backbone for marine robotics FAIR data protocols feeds into the broader ideal of creating interdisciplinary and international science. This concept is strongly supported in the European Open Science Cloud (EOSC) (EOSC Future H2020 Grant) resource, which, especially for Earth Science, is extended by projects such as RELIANCE (RELIANCE Grant H2020). In fact, RELIANCE aspires to manage the research lifecycle using a novel tool: research objects.

2. Outcome

2.1. Domain agnostic guidelines

A field deployment using a novel marine robot produces a set of data that can be divided into two macro groups, environmental and robotic. If the deployment has a robotic focus we tend to neglect the environmental data, consequently hindering its quality; on the other hand if the deployment has an observational meaning the robotic data are not stored, and if they are this happens following internal practices, which are highly variable depending on the research group. We propose a unified, yet, customised approach, as autonomous as possible, to value both data groups, rendering them

mutually validating.During a given data campaign, with an innovative or traditional robotic platform, the descriptive metadata, registered for the environmental and for the device data are the same. The differences lay in the content of the metadata, i.e. data convention applied, name of the collecting platform operator and so on, but the fields populating the descriptive metadata are the same (**Figure 1**). Hence, being descriptive metadata domain agnostic, marine robotics can uptake the standards and conventions already available, for instance ISO19115. Nowadays the most reliable and stable format to share and store data is NetCDF - Network Common Data Form (Network Common Data Form, UCAR).

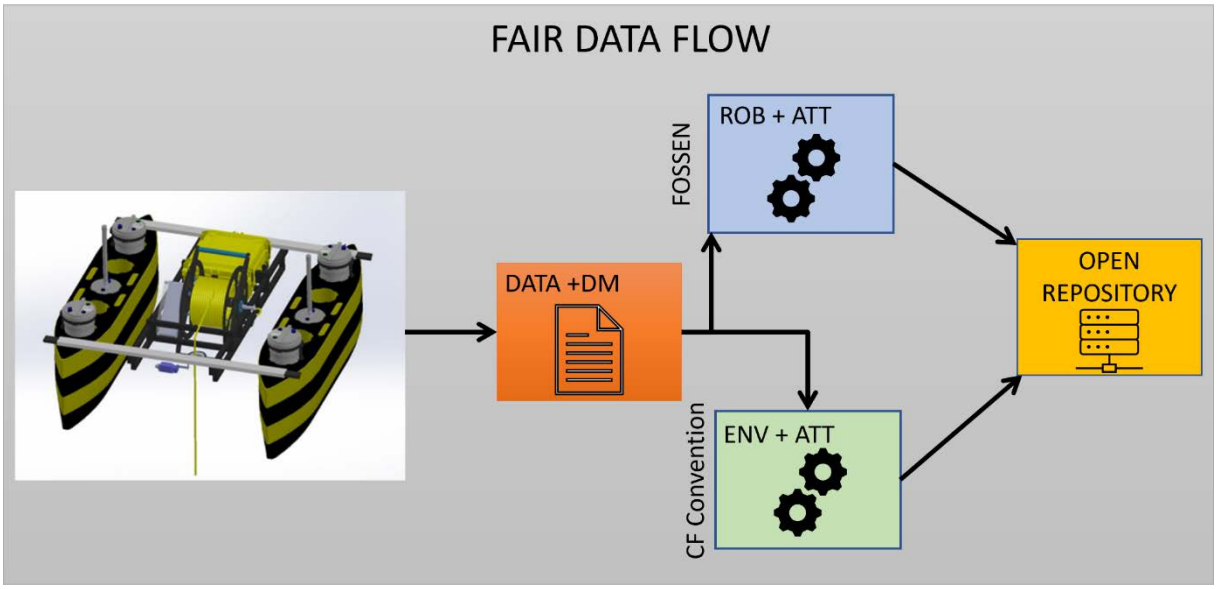

Figure 1. FAIR Data flow, from collection with a novel marine robotic device (SWAMP). Initially descriptive metadata (domain agnostic) are compiled for the whole dataset, then the variable specific metadata are compiled separately for robotic and environmental data, referring respectively to Fossen based and CF shared vocabulary. The joint dataset can then converge in a standardised format - e.g. NetCDF - and shared in an open repository.

2.2. Domain specific guidelines

The metadata describing the measured variables in Earth sciences are set, coming from a long tradition. It is also not uncommon to have to homogenise data coming from very different platforms, especially when studying long time series, which strongly encourages the implementation of data standards. In line with it, the CF (Climate and Forecast) (Hassell et al., 2017) Metadata Conventions offers a set of guidelines and, more importantly, a shared vocabulary to store and describe variables following the FAIR principles. Hence, when it comes to populating a NetCDF file following FAIR data principle the steps to follow are clear. Trying to do the same with marine robotic data is not straightforward. In fact, to the best of our knowledge, the only reference document is the book Guidance and Control of Ocean Vehicles by Fossen (1994). Merging the Fossen nomenclature and the data policies coming from Earth sciences, **Figure 1** shows the workflow to undertake to produce a combined FAIR dataset during a field campaign employing a novel robotic platform, such as SWAMP (Shallow Water Autonomous Multipurpose Platform) (Odetti et al., 2020).

3. Discussion

As shown, the need for metadata in marine robotics is urgent and an absolute necessity for open data. The work presented here is a step towards standard holistic data protocols joining environmental and marine robotic data. However, it needs to be shared and disseminated among the marine robotic community. One possibility to be explored is through the IEEE Oceanic Engineering Society (OES) Standards Committee. This Committee can be a good avenue to involve a broader community and make sure that any metadata standardization efforts are commonly agreed and adopted. By doing so, the findability and interoperability of marine robotics datasets will improve and open data will become more valuable. With an explosion in marine robotics applications and in the number of research groups advancing the state of the art and collecting large amounts of data, this need becomes even more important. Conveying the recommendations hereby presented will ultimately enhance the FAIR status of cutting edge marine robotics datasets.

Acknowledgements: This work has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101060395 - Montenegrin Centre for Underwater Sensor Networks (MONUSEN). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. This work was supported by the European Regional Development Fund through the Interreg Italy-Croatia InnovaMare project (Partnership ID 10248782). The field campaigns with marine robots would not be possible without the extremely valuable work of the CNR-INM's technical team, thank you Giorgio Bruzzone, Edoardo Spirandelli and Mauro Giacopelli.

References

Bernardi M, Hosking B, Petrioli C, Bett, J. B., Jones, AI Huvenne, V., Marlow, R., Furlong, M., McPhail, S., Munafò, A. (2022). AURORA, a multi-sensor dataset for robotic ocean exploration. The International Journal of Robotics Research 41(5): 461-469. https://doi.org/10.1177/02783649221078612

Burgelman, J-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., Schouppe, M. (2019). Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. Frontiers in Big Data 2: 43. https://doi.org/10.3389/fdata.2019.00043

DataCite Schema (2021). Available at https://schema.datacite.org/

ISO (2022). Date and time format. ISO 8601. Available at https://www.iso.org/iso-8601-date-and-time-format.html

DCMI (2022). Dublin core. Available at https://www.dublincore.org/specifications/dublin-core/

EOSC Future - Advancing European research through Open Science. H2020. Grant agreement ID: 101017536. https://doi.org/10.3030/101017536. Coordinated by ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS. EC signature date: July 5, 2021. Record number: 236671.

Fossen, T. I., (1994). Guidance and Control of Ocean Vehicles 494.

Fritz, S., See, L., Carlson, T. et al. (2019). Citizen science and the United Nations Sustainable Development Goals. Nature Sustainability 2: 922–930. https://doi.org/10.1038/s41893-019-0390-3

Hassell, D., Gregory, J., Blower, J., Lawrence, B. N., Taylor, K. E. (2017) A data model of the Climate and Forecast metadata conventions (CF-1.6) with a software implementation (cf-python v2.1), Geoscientific Model Development 10: 4619–4646, https://doi.org/10.5194/gmd-10-4619-2017

Hörstmann, C., Buttigieg, P.L., Simpson, P., Pearlman, J., Karstensen, J. Waite, A.M. (2020) Towards a Best Practice for Developing Best Practices in Ocean Observation (BP4BP): Supporting Methodological Evolution through Actionable Documentation. Paris, France, UNESCO, 33. http://dx.doi.org/10.25607/OBP-781

ISO19115 Geographic information — Metadata — Part 1: Fundamentals. Available at https://www.iso.org/standard/53798.html

Marine Regions. Managed by Flanders Marine Institute. Available at https://www.marineregions.org/. Accessed December 7, 2022.

NATO, STANDARD ANEP-87, (2017). Digital Underwater Signalling Standard for Network Node Discovery & Interoperability. Edition A Version 1. NATO STANDARDIZATION OFFICE (NSO). Available at https://nso.nato.int/nso/nsdd/main/standards?search=ANEP-87

Network Common Data Form (NetCDF) by UCAR/Unidata. https://doi.org/10.5065/D6H70CW6

NORSOK Standard, U-102 (2016). Remotely Operated Vehicle (ROV) Services.

Odetti, A., Bruzzone, G., Altosole, M., Viviani, M., Caccia, M. (2020). SWAMP, an Autonomous Surface Vehicle expressly designed for extremely shallow waters, Ocean Engineering 216: 108205. https://doi.org/10.1016/j.oceaneng.2020.108205

Ramachandran, R., Bugbee, K., Murphy, K. (2021). From open data to open science. Earth and Space Science 8: e2020EA001562. https://doi.org/10.1029/2020EA001562

REsearch LIfecycle mAnagemeNt for Earth Science Communities and CopErnicus users in EOSC (RELIANCE) - Extending the research-enabling capabilities of the European Open Science Cloud. H2020. Grant agreement ID: 101017501. https://doi.org/10.3030/101017501. Coordinated by INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK. EC signature date: December 1, 2020. Record number: 232563

The NERC Vocabulary Server (NVS) (2021). Available at https://vocab.nerc.ac.uk/. Accessed December 7, 2022.

Waldmann, C., O'Reilly, T., Jiang, F., Ferreira, F., Xiang, X., McCummins, R., Pallayilline, V., (2021). About the Value of Standards for Ocean Technology. OCEANS 2021: San Diego – Porto, 1-5. https://doi.org/10.23919/OCEANS44145.2021.9705984

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3: 160018. https://doi.org/10.1038/sdata.2016.18

DIGITAL INDIA MISSION: REVOLUTIONISING OPEN DATA MECHANISMS SINCE 2015

Aditi Basu[1*]

[1]Independent researcher, India

*correspondence E-mail: aditibloyolajsr@gmail.com

Keywords: Covid-19; electronic transactions; Open Data; digital revolution

1. Introduction

Open Data Mechanism in India involves the usage of Information and Communication Technology (ICT) in governance. It adjudicates for SMART governance which implies Simple, Moral, Accountable, Responsive and Transparent, better known as E-Governance for adopting a citizen-centric approach, improving the quality and delivery of public services. E-Governance has proved to be cost-effective, inclusive, holistic, easy and reliable. The National E-Governance Plan (NeGP) was approved in 2006 with short-term goals, the scope of which were further expanded in 2011 with a lot of schemes and initiatives in Health, Education, Public Distribution System (PDS), Security, Human Resources Development, Enterprise Architecture, Banks and Post-Offices.

The paper shall explain the paradigm shift in India's electronic and digital revolution in its open and transparent data mechanism since 1990s. It will also discuss the schemes and initiatives of the Indian government both at the national and regional level to ensure transparency and accountability. It shall explain, in detail, the achievements and challenges of the Digital India Mission in the past seven years in general, and the different initiatives in particular. Special attention shall be paid to its immense contribution in empowering the world's largest democracy. Finally , the paper will highlight the elaborate revolution that the Indian government has made globally and its contribution towards tackling the Covid-19 Pandemic.

2. Background

Launched by Prime Minister Shri Narendra Modi in 2015, the Digital India campaign, the Ministry of Electronics and Information Technology (MeitY), ensures the accessibility of government's services to the citizens on an online platform, improving their digital literacy and enhancing broadband connectivity (Ministry of electronics and Information Technology, Government of India, 2022). It also seeks to make the Indian economy a knowledge-centric and a digitally empowered one. The campaign involves Initiatives like Bharatmala, BharatNet, Make in India, Startup India, Standup India and Sagarmala.

The first Open Data Mechanism of India was launched in 2012 named as the National Data Sharing and Accessibility Policy (NDSAP) which formally conferred on the National Informatics Centre the responsibility of the Open Government Data Platform. The Open Government Data (OGD)

Platform of India is the sixth pillar (Information for All) under the Digital India Mission for the Central government departments and ministries to officially publish and update their data, statistics, catalogues, documents and applications in an open format for ensuring a responsible, accessible, innovative and transparent functioning of government within its framework through workshops, hackathons, Work-flow based Data Management systems, portals and application softwares. It also enables the citizens and stakeholders to express their views, suggestions and opinions on the different government policies and initiatives, conduct researches on them and seek clarification and explanation from the nodal officers at different levels of the government. A separate Community Portal (https://community.data.gov.in) acts as a platform to exchange ideas and share knowledge through blogs. The second phase of OGD, that is, OGD 2.0 is operative under the NITI Aayog as the National Data and Analytics Platform and under the Ministry of Housing and Urban Affairs (MoHUA) as the India Urban Data Exchange. In future, the improved environment for Open Data shall act as a critical enabler for boosting the nation's Artificial Intelligence Mechanism (Carmeli, 2021). The NDSAP has enabled wider accessibility of data and it is still working for the enhancement of wider data usability. The MeitY has published the Draft India Data Accessibility & Use Policy 2022 for the standardisation and establishment of metadata management framework for the improvement of quality and accessibility of non-personal government datasets which was further replaced by the Draft National Data Governance Framework Policy (Vats, 2022). After the launch of NDSAP, many academic circles, higher education institutions, governmental and civil society organisations have launched their own Open Data platforms, for example, the India Data Portal by the Indian School of Business and the Open Budgets India portal by the Centre for Budget and Governance Accountability. The use of Artificial Intelligence (AI) is prevalent in India's education, healthcare and agricultural spheres, for example, the Shodhganga contains all the Doctoral theses of the research scholars.

3. Platforms facilitating Open Data Mechanism

Rural and urban areas of India have been provided with broadband connectivity under the Broadband Highways initiative and Universal Access to Mobile Connectivity initiative. Common Service Centres and Post Offices for providing internet accessibility and delivering government and business services in Panchayats have been transformed into multi-service centres. The governance process in public and private offices, academic institutions, industries and banking institutions have been made transparent with the use of IT by online tracking, electronic database, online grievance redressal, online repositories, speedy service delivery and interface between the departments (Press Information Bureau, 2022). The mechanism of biometric attendance is mandatory in all institutions in India that provides a transparency in the maintenance of records of the absence and presence of the concerned people.

The major apps that have brought a transformation in India's Open Data Mechanisms are:

1. Startup India- Announced in 2015, the app serves as a platform to allow the startups develop a network in accessing free tools and resources and knowledge-sharing through workshops and programmes.
2. MyGov- The platform actively engages citizen participation in governance and adopts a bottom-up approach for strengthening the democracy.

3. UMANG serves as a single window portal/ application software for providing services to the citizens at all levels of government.
4. Aarogya Setu is a mobile application for providing essential and immediate health services to the people in combating Covid-19.
5. BHIM App is a mobile payment application software been created by the National Payments Corporation of India on the Unified Payments Interface (UPI) in the year 2016. It has helped in the promotion of cashless transactions by the facilitation of direct e-payments.
6. ENAM portal (Electronic National Agriculture Market) is a trading website that is supported by the Indian government to connect small grocery shops to develop a national market for agricultural products by making it easier for the farmers to remove the trade barriers and receive information about the prices, demand and supplies of their respective products on their smartphones.
7. Mausam App has been built by the Ministry of Earth Sciences, Government of India to observe and forecast weather changes and obtain radar photos.
8. Agrimarket Mobile App provides information about the prices of the crops to the farmers within 50 kilometres of their own device.
9. E-Kranti Mission has many Mission Mode Projects for the use of technology in education (Massive Open Online Courses), healthcare, farmers (online payments), mobile based energy services and disaster- related services, mobile banking, Micro-ATM Programme and post offices, online hearing of court proceedings and cyber security.
10. The PM SVANidhi App (Pradhan Mantri Street Vendors' Atmanirbhar Nidhi), launched by the Ministry of Housing and Urban Affairs, Government of India, incentivises digital transaction amongst the street vendors by providing them information about the affordable loans at cheaper interest rates by the governments to develop their businesses.
11. The Jaldoot App has been launched by the Ministry of Rural Development for measuring the groundwater levels twice a year, that is, once in pre-monsoon months and the other in post-monsoon months (NewsOnAIR, 2022).
12. Apps like the Electronic- Know Your Customer (e-KYC), the Electronic Document Storage System (DigiLocker) and the Electronic Signature System (eSign) were introduced to help businesses streamline their operations.

4. International Cooperation on Open Data Mechanisms

International cooperation in global digital policy between India and other countries is a potential research area. Recently, the Hon'ble Finance Minister of India Smt. Nirmala Sitharaman said that India's Open Network for Digital Commerce has revolutionised the retail and manufacturing sector and envisioned plenty of scope for collaboration in the digital sector between India and the United States. The federal government was engaged with foreign investors to ease rules for more investments. She also said the country was confident of handling the challenge of high inflation and that its economic revival was driven by government reforms.

Cooperation in global digital policy is considered one of the most promising fields in the strategic partnership between India and the European Union (EU). However, profound differences are apparent in terms of implementation, for example with regard to data protection, competences of security authorities and the future global digital order. Meanwhile, similar problems are being addressed in the EU's negotiations with the US on digital trade issues. Possible compromises there

could also form components of an understanding with India. Shared democratic values are consistently referred to as a justification for efforts to strengthen Europe's cooperation with India. In their Roadmap 2025, India and the EU affirm their interest in promoting an "open, free, stable and secure cyber-space" and fighting cybercrime. But the road to this goal is proving to be rocky (Voelson and Wagner, 2022). They are willing to cooperate in areas like an open, free, stable and secure cyberspace and cyber security, ICT cooperation under the India-EU Joint Working Group on sustainable digital infrastructure, services, norms and regulatory frameworks, ensuring interoperability of networks, promoting international standards of data protection, quantum computing, artificial intelligence, agritech, healthtech and blockchain.

Recently India, the EU and nine countries signed the "Joint Declaration on Privacy and the Protection of Personal Data: Strengthening Trust in the Digital Environment" for fostering international cooperation to promote high data protection and privacy standards based on certain core elements increasingly shared across the Indo-Pacific region, Europe and beyond. The European Union, Australia, Comoros, India, Japan, Mauritius, New Zealand, South Korea, Singapore and Sri Lanka said rapid technological developments, in particular in information and digital technologies, have brought benefits for their economies and societies, as well as new challenges for privacy and the protection of personal data.

5. Achievements and Challenges

One of the biggest achievements of India's Open Data Mechanism is that it has led to greater accessibility and transparency of government's digital services to the citizens. An umbrella programme covering innumerable initiatives and schemes under the MeitY, the UPI initiative has flourished the businesses of street vendors to business tycoons. It has helped in connecting the citizens with more than 300 public services. The Digital India campaign with Digital Village brought into reality the concept of a traditional India into a modern, tech-savvy one. Speed was achieved in the affairs of the government and private businesses. Public information has become readily accessible in your fingers. For this, it has provided high-quality internet connectivity in 2.5 lakh villages and lowered the data tariffs (Chaturvedi, 2021). The role of CivicDataLab in the skill enhancement of the government officials in finance and accounting sectors is commendable.

The performance of the Open Data Mechanism in India in the fields of healthcare, trade and commerce, banking, agriculture, education and research, smart cities projects and mobility has been splendid. A more synchronised database management between the central and state governments is needed since India has a vast diverse population. For a diverse country, data collection, editing, tabulation and making it digitally transparent becomes a Herculean task. More hackathons, educational and training programmes, capacity-building workshops with skilled data contributors and resource persons are needed for the spread of digital literacy among the citizens and also for the upgradation of the existing database to higher quality and higher value ones (Vats, 2022). The collective participation of civil society, government and industry is needed for a more dynamic, spontaneous and open artificial intelligence mechanism in governance and benefitting the Indian corporations and startups. An Open Data legislation is the need of the hour that lends uniformity and robustness for the careful selection of datasets and their expansion to state governments. More involvement of researchers, academia, IT professionals and data managers is required to meet the rising demands of Open Data. A uniform Personal Data Protection legislation is also needed for

safeguarding the privacy, security and rights of individuals in case of anonymised data (Neufeld, 2021).

References

Ajaykumar, S. (2022). Democratising the Internet: Landscape of digital markets and its impact. Observer Research Foundation (ORF). Available at https://www.google.com/amp/s/www.orfonline.org/expert-speak/democratising-the-internet/

Aswath, K. (2022). How digitalisation can revolutionise your B2B e-commerce business. The Times of India. Available at https://timesofindia.indiatimes.com/blogs/voices/how-digitalisation-can-revolutionise-your-b2b-e-commerce-business/

Bagchi, I. (2021). New Quad takes shape with focus on tech, trade, big data. The Times of India. Available at https://timesofindia.indiatimes.com/india/new-quad-takes-shape-with-focus-on-tech-trade-big-data/articleshow/87147685.cms

Carmeli, S. (2021). To Unlock the Potential of Open Government Data, India Needs New, Comprehensive Legislation. The Bastion. Available at https://thebastion.co.in/politics-and/tech/to-unlock-the-potential-of-open-government-data-india-needs-new-comprehensive-legislation/

Chaturvedi, S. (2022). Super Six: 6 years of Digital India. Financial Express. Available at https://www.financialexpress.com/life/technology-super-six-6-rears-of-digital-india-2281881/

Dani, R. (2022). Digital India is a global success story. The Times of India. Available at https://timesofindia.indiatimes.com/blogs/thedanispost/digital-india-is-a-global-success-story/

Digital bridge to growth; Know about major apps launched by Central Government. (2022). News on Air. Available at https://newsonair.com/2022/09/29/digital-bridge-to-growth-know-about-major-apps-launched-by-central-government/

Digital India. Official website of Ministry of Electronics and Information Technology, Government of India. Available at https://www.digitalindia.gov.in/content/programme-pillars

Digital India Conference of state IT Ministers held along with launch of 5G (2022). ANI. Available at https://www.aninews.in/news/national/general-news/digital-india-conference-of-state-it-ministers-held-along-with-launch-of-5g20221004065032/

E-Kranti. Digital India, Department of Electronics and Information Technology, Government of India. Available at https://www.digitalindia.gov.in/content/ekranti#:~:text=e%2DKranti%20(Transforminge%2DGovernanceforTransforming,the%20Union%20Cabinet%20on%2025.03

EU, India, 8 other countries call for international cooperation on personal data protection. (2022). The Indian Express. Available at https://indianexpress.com/article/india/eu-india-international-cooperation-personal-data-protection-7831126/lite/

India's digital revolution offers investment opportunities to US, says FM Nirmala Sitharaman (2022). The Indian Express.
Available at https://www.google.com/amp/s/indianexpress.com/article/business/economy/indias-digital-revolution-offers-investment-opportunities-to-us-says-fm-nirmala-sitharaman-8135937/lite/

India-ASEAN Digital Work Plan 2022 approved at 2nd ASEAN Digital Ministers (ADGMIN) meeting. Press Information Bureau, Ministry of Communications, Government of India. Available at https://pib.gov.in/newsite/PrintRelease.aspx?relid=231114

India-EU Strategic Partnership: A Roadmap to 2025. Official website of the Ministry of External Affairs, Government of India. Available at https://www.mea.gov.in/bilateral-documents.htm?dtl/32828/IndiaEU_Strategic_Partnership_A_Roadmap_to_2025

India to lead draft plan for cybersecurity in BIMSTEC: Official. (2022). The Hindustan Times. Available at https://www.hindustantimes.com/india-news/india-to-lead-draft-plan-for-cybersecurity-in-bimstecofficial-101657995791995.html

Jain, S. (2022). The G-20 Digital Economy Agenda for India. Observer Research Foundation. Available at https://www.orfonline.org/research/the-g20-digital-economy-agenda-for-india/

MeitY holds Brainstorming with Researchers & Startups to shape strategy for Digital India BHASHINI - India's AI led language translation platform. Press Information Bureau, Official website of Ministry of Electronics and

Information Technology, Government of India. Available at https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1827997

Mishra, M. (2022). To better track PLI claims, Govt. floats digital platforms for data sharing. The Indian Express. Available at https://indianexpress.com/article/india/to-better-track-pli-claims-govt-floats-digital-platforms-for-data-sharing-8183549/

National E-Governance Plan. Official website of Ministry of Electronics and Information Technology, Government of India. Available at https://www.meity.gov.in/divisions/national-e-governance-plan

Neufeld, S. (2021). Deploying open government data for AI-enabled public interest technologies. Observer Research Foundation (ORF). Available at https://www.orfonline.org/expert-speak/ai-driven-public-interest-technologies-employing-open-government-data-achieve/

Nine pillars of Digital India. Press Information Bureau, Vikaspedia. Available at https://vikaspedia.in/e-governance/digital-india/nine-pillars-of-digital-india

Open Government Data (OGD) Platform, India. National Informatics Centre (NIC). Available at https://www.nic.in/products/open-government-data-ogd-platform-india/

Pareek, S. (2022). 12 Projects You Should Know About Under the Digital India Initiative. The Better India. Available at https://www.thebetterindia.com/27331/12-projects-you-should-know-about-under-the-digital-india-initiative/

Parihar, I. (2015). How is open data changing India? World Economic Forum (WEF). Available at https://www.weforum.org/agenda/2015/02/how-is-open-data-changing-india/

Prasad, V. (2022). Why India's Data Protection Bill Should Follow the EU Model, Not the China-Russia One. The Wire. Available at https://thewire.in/rights/data-protection-bill-localisation-eu-china-russia

PM Modi lays roadmap for 'Digital India' expansion to BIMSTEC member nations, TTF in Colombo soon. (2022). The Economic Times. Available at https://government.economictimes.indiatimes.com/news/digital-india/pm-modi-lays-roadmap-for-digital-india-expansion-to-bimstec-member-nations-ttf-in-colombo-soon/92221055?redirect=1

Railways promoting Digital India initiative to encourage digital transactions (2022). LiveMint. Available at https://www.livemint.com/news/india/indian-railways-promoting-digital-india-initiative-to-encourage-digital-transactions-11663890523549.html

Vats, A. (2022). A decade into India's open government data journey. Observer Research Foundation (ORF). Available at https://www.orfonline.org/expert-speak/a-decade-into-indias-open-government-data-journey/?amp

Voelson, D., Wagner, C. (2022). India as an Ambivalent Partner in Global Digital Policy. Stiftung Wissenschaft und Politik German Institute for International and Security Affairs.

Vipra, J., Dhruv, S. (2022). Access to do Digital Business in BRICS countries. VIDHI Centre for Legal Policy. Available at https://vidhilegalpolicy.in/blog/access-to-do-digital-business-in-brics-countries/

SELECTION OF ITS SOLUTIONS FOR IMPROVING THE QUALITY OF THE TRAFFIC SYSTEM IN THE CITY OF ZAGREB BASED ON USER REQUIREMENTS

Antun Mandić[1], Antonio Bubnić[1], **Marko Matić[1]**, Filip Bišćan[1], **Andre Garašić[1*]**, Miroslav Vujić[1]

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

*correspondence E-mail: andre.garasic@gmail.com

Keywords: Intelligent Transportation System (ITS); centre for data processing and traffic management; user requirements; passenger information; Public Transport (PT)

1. Introduction

The quality of the transport system in cities is significantly improved using intelligent transport systems (ITS), which implies the use of advanced telematic solutions and other technologies. ITS can be defined as a holistic and information communication superstructure of the classic traffic and transport system, which can achieve a significant improvement in traffic flow, more efficient transport, increase in traffic safety, comfort and protection of passengers, less environmental pollution, etc. (Bošnjak, 2006). In the initial reference model for the ITS sector, 11 functional areas and 32 services were defined. Some of these functional areas are passenger information, traffic management, driver assistance and vehicle control.

The implementation of intelligent solutions makes it possible to fulfil user requirements for a given system, such as the optimization of adaptive traffic light systems, the traffic management centre, the public transport (PT) passenger information system, giving unconditional priority to urgent services, the automatic payment system for entrance to the city centre, and the park & ride system.

Sensors and detectors are key technical components of ITS that enable data collection from source systems. The sensor represents the input and output of the system that can react to light, heat, pressure, electric or magnetic field, and gas so that it produces a signal about the state of the medium where it is located. The output is usually an electrical signal that is further processed and transmitted to the control part of the system (Ezgeta, 2018).

All collected data can be processed at the same level and provide a solution based on the current state. For the optimal operation of the entire system, the centralization of the data would lead to a greater impact on the entire transport system. Some obstacles faced by this topic are that such a centre does not exist, different stakeholders in terms of disciplines are needed to run a centralized system, separate companies that own data are not willing to share with each other and data processing itself would be challenging due to the amount and different formats of records.

2. Survey results and analysis

In the paper, a survey was conducted to find out from the users what would be the ideal solution to improve the traffic system of the city of Zagreb. In total, there were about 150 respondents in the area of Zagreb and its surroundings. Originally, the survey collected data about the person filling out the survey. This includes information about the person's age, gender, and employment. In the survey, participants were asked to describe which form of transport they use and how often they travel to Zagreb. Knowledge of various terms such as Intelligent Transport Systems, Old Town Entrance Fee, Park & Ride, Main Traffic Centre was especially fulfilling. In the next phase of the survey, people expressed their opinion on the current state of the transport system, satisfaction with the transport system in the city of Zagreb in the centre and surroundings, whether public city transport provides users with reliable information when travelling, the level of satisfaction with the work of emergency services and whether it is necessary to introduce the main traffic control centre.

The conducted survey revealed that a large proportion of people are between 20 and 40 years old. By examining users about their gender, it was determined that 55.62% of respondents were men, while 44.37% were women. The largest share of respondents use public transport and a personal vehicle as the primary means of transport, while transport by train and other means is poorly represented. According to the respondents, the quality of the traffic system of the city of Zagreb was rated with an average rating of 3.13. when choosing a grade, there is a small concentration of excellent (5) and very poor grades (1). The quality of traffic management within the core of the city of Zagreb was also rated with an average rating of 3.04. Among the respondents, 88.08% are of the opinion that the traffic system in the city of Zagreb needs to be reorganized. Public city transport was evaluated with an average score of 3.07, while the reliability of informing passengers in transport is not considered sufficiently reliable by 56.60% of respondents. In the context of public city transport, the vast majority (94.70%) believe that it is necessary to improve the tram and bus system, and they also believe that the quality of the displayed real time data should be improved. The majority opinion of the respondents shows that right-of-way should be added to public city transport, especially in the old city core. The quality rating of emergency services is 3.43, while their response to incidents is considered acceptable. 52.98% of respondents knew the concept of the main traffic centre, and 86.75% of them believed that it is needed in the city of Zagreb.

3. Possible ITS solution

Considering the collected data, a solution is found that is ideal for improving the traffic system of the city of Zagreb. Using the analytical hierarchy process method, it is found that the centre for data processing and traffic management is the best in all aspects compared to other solutions.

The centre for data processing and traffic management is a facility that is fully cooperative with the traffic infrastructure and connected by numerous cameras, sensors, detectors and other communication methods. Data processing performed within the control centre is in real-time, which enables easier management of the entire traffic network and insight into possible traffic congestion and incidents. The monitoring centre monitors almost the entire traffic network and sends users the necessary information and redirects them as needed. The control centre has the

ability to communicate with users via radio connection, electronic signs or via mobile devices that must be connected to the Internet or Bluetooth in advance.

The entire system consists of a input subsystem (entire sensors, detectors, inductive loops, video surveillance), a communication subsystem (Bluetooth connection, wireless 4G, radio connection) and an infrastructure subsystem (objects closely related to the control centre) (Vidaković, 2017).

4. Data collection processes

One of the main systems applied in the city of Zagreb through which data would be collected on the state of the part of the traffic system in which the device is located are: electronic bulletin boards, ticket registration, smart garages, interactive electronic panels, pedestrian buttons and navigation systems - GPS navigation (Bošnjak, 2006).

Data on public city transport traffic would be collected in the data processing and traffic management centre from ticket registration devices, electronic billboards, interactive electronic panels, and GPS navigation of individual public city transport entities.

Electronic billboards are located at terminals, tram and bus stations where each billboard has the option of displaying passenger information about the current traffic situation, significant changes in the timetable and cancellation of PT lines (United States Department of Transportation, 2022). Information that can be found on the screen includes the arrival and departure times of a particular bus or tram line, the identification number of the next PT vehicle and other information for that line. The terminal would coordinate the same data with the centre for data processing and traffic management using some of the wireless communication methods.

Interactive electronic panels are computer circuits with appropriate software that enables insight into timetables, possible delays, the current position of PT vehicles, a map with stops and terminals, and the annexation of passengers. At the same time, they enable the display of information to the carrier, ticket prices and weather information. The most common places where interactive panels are installed are terminals, but also in infrastructure assemblies at PT stops or as an independent object. The data it would display would be synchronized with the data processing and traffic management centre.

The GPS coordinates of individual traffic entities of public city transport would also be available to the centre for data processing and traffic management and, in combination with other data from various sensors and detectors, would show the real time state of the traffic network.

Smart public garages are a special form of garages that, first of all, provide users with information about the number of free spaces, and in advanced versions, users are guided to the parking space. In addition to the technology used in ordinary garages (ramp, card printer), "smart" has been added, which has a known number of free spaces in the garage and compares them with the entrance and exit of vehicles and shows everything on the LCD display when entering the garage. The distance sensor scans the distance and reports to the central computer whether the place is free or occupied. By adding a camera at the entrance and exit, it is possible to scan the registration to make it easier to miss the vehicle when leaving the garage. Navigation through the garage is made possible by an LCD display that directs certain cars with an arrow to a free space. This type of system generally

reduces waiting times and crowds inside the garage, but it could also be useful in a wider traffic area. Using the data of the smart garage system, the traffic management centre could optimize the load on garages at the level of the city of Zagreb, thereby redirecting a certain number of vehicles to zones with less traffic load.

5. Conclusion

Having that said, using devices for data collection at the intersection, such as cameras, pedestrian button counters and other sensors and detectors, would enable the data processing and traffic management centre to have a better overview of the traffic situation in the city of Zagreb. Access to this data would enable traffic to be adjusted to the actual situation. In addition, they would thereby contribute to the reduction of the environmental impact of traffic on the environment, increasing both safety and passenger satisfaction.

After the definition of the present state of traffic management in urban areas, it is possible to conclude that such an element is needed for the optimisation and regulation of urban traffic networks. Because a large amount of data can be gathered from the traffic network, it is essential to open the collected datasets for wide usage for all involved stakeholders and for end users. Traffic data on this higher level of openness provides the basis for the development and implementation of advanced ITS solutions designed for the improvement of the urban traffic network. Future research will be focused on the allocation of relevant traffic data which must be opened and to what extent. Also, with the opening of allocated data, which ITS solutions in urban areas can be upgraded and implemented for a better urban traffic network quality.

Acknowledgements: The authors acknowledge the financial support from TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Bošnjak, I. (2006). Inteligentni transportni sustavi -ITS 1. Faculty of Transport and Traffic Sciences, University of Zagreb, Zagreb.

Ezgeta, D. (2018). Inteligentni transportni sustavi. Faculty of Traffic and Communications, University of Sarajevo, Sarajevo.

U.S. Department of Transportation (2020). Benefits of intelligent transportaion systems (ITS). Available at https://www.its.dot.gov/factsheets/pdf/Benefits_FactSheet.pdf. Accessed June 2, 2020.

Vidaković, F. (2017). Concpet of development center to monitor and manage traffic in cities (in Croatian). Graduation thesis. Faculty of Transport and Traffic Sciences, University of Zagreb, Zagreb.

SESSION V

LEGAL CHALLENGES TO OPEN DATA

OPEN DATA, DATA MINING AND PERSONAL DATA IN LAW ENFORCEMENT ENVIRONMENT

Mišo Mudrić[1]

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

*correspondence E-mail: miso.mudric@pravo.hr

Keywords: Facial recognition; Law enforcement bodies; LED Directive; Draft EU AI Act; CCTV surveillance

In recent years, the law enforcement bodies have increasingly begun to use the facial recognition technology as a tool for identifying suspects and solving crimes. This technology is based on the use of algorithms that can analyze and match facial features to a database of known individuals. One of the main advantages of using facial recognition technology is that it can help to identify suspects quickly and accurately. This can be especially useful in cases where traditional methods of identification, such as fingerprint analysis or DNA testing, are not available or are not effective. Facial recognition technology can be used to monitor large crowds, such as at public events or in high-crime areas, to identify potential suspects or persons of interest. Another benefit of using facial recognition technology is that it can help to reduce the number of false arrests and wrongful convictions. This is because the technology is based on objective criteria and is not subject to human bias or error. Additionally, facial recognition technology can be used to search for missing persons, such as children or elderly individuals, and can assist in the identification of human remains.

Closed-circuit television (CCTV) cameras have become a common sight in many cities around the world, and law enforcement bodies have begun to use them as a tool for mass surveillance. This technology is based on the use of cameras that can capture images and videos of individuals in public spaces and transmit them to a central monitoring station. One of the main advantages of using CCTV cameras for mass surveillance is that they can help to deter crime. This is because the presence of cameras can make individuals think twice before committing a crime, as they know that they might be caught on camera. CCTV cameras can be used to monitor high-crime areas, such as city centers or public transportation systems, in order to identify and track suspects. Additional benefit of using CCTV cameras is that they can help to increase the efficiency of law enforcement. Cameras can provide real-time footage that can be used to track suspects and assist in the identification of suspects. CCTV cameras can also be used to monitor large crowds, such as at public events, to ensure public safety and to identify potential suspects or persons of interest.

There are, however, potential downsides to using facial recognition technology and CCTV cameras. One of the main concerns is that they may be used to violate individual privacy rights. There are also concerns that the technology could be used to target certain groups, such as people of color

or those with certain physical characteristics. Furthermore, the accuracy of facial recognition technology can be affected by factors such as lighting, angle, and facial expressions, which may lead to false positive or false negative identifications. The use of CCTV cameras for mass surveillance can lead to a feeling of being constantly monitored and can have a negative impact on mental health. In addition, the presence of CCTV cameras can lead to a chilling effect on free speech and freedom of assembly.

Facial recognition technology has the potential to be a valuable tool for law enforcement bodies in their efforts to identify suspects and solve crimes. It is critical to ensure that the technology is used in a way that respects individual privacy rights and does not lead to discrimination or bias. Measures must be taken to ensure the accuracy of the technology and to minimize the risk of false identifications, to ensure the protection of individuals' rights and to minimize the negative impact of surveillance on mental health and civil liberties. There should be a clear and transparent legal framework for the use of CCTV cameras for mass surveillance and a system of oversight and accountability to ensure that the technology is used in a responsible and ethical manner.

Among other relevant regulatory instruments, two documents stand out. The EU Law Enforcement Directive, also known as the Directive on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters, is a piece of legislation adopted by the European Union in 2016. The Directive aims to protect personal data processed by EU law enforcement bodies in the context of preventing, investigating, detecting, or prosecuting criminal offenses. One of the main goals of the Directive is to ensure a high level of protection for individuals' personal data in the context of law enforcement activities. To achieve this, the Directive sets out a series of rules and principles for the processing of personal data, including the need for a legal basis for processing, transparency and fairness, and the right to access, rectify, and delete personal data.

The EU Draft Artificial Intelligence Act is a proposed piece of legislation that aims to regulate the use of artificial intelligence (AI) in the European Union. The Act, which is currently in the draft stage, seeks to ensure the safety and security of AI systems, as well as to protect individuals' rights and freedoms in relation to their use. One of the main goals of the Act is to ensure the safety and security of AI systems. This includes requirements for transparency, traceability, and accountability, as well as measures to ensure that AI systems are robust, reliable, and trustworthy. Additionally, the Act includes provisions for the testing and certification of AI systems, as well as for the reporting of incidents and accidents.

Real-time mass surveillance is an area that the EU Draft Artificial Intelligence Act is also addressing. The Act aims to regulate the use of AI systems for the purpose of mass surveillance, including the use of facial recognition technology, and to ensure that such systems are used in a way that respects individuals' rights and freedoms. This includes the requirement for a legal basis for the processing of data, the need for transparency and fairness, and the right to access, rectify, and delete data.

RIGHTS IN DATABASES AND OPEN DATA POLICIES – CONVERGENCE OR CONFLICT?

Romana Matanovac Vučković[1]

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

*correspondence E-mail: rmatanov@pravo.hr

Keywords: open data; copyiright in databases; sui generis database right; Directive 96/9/EC; Directive (EU) 2019/1024; Directive (EU) 790/2019

Directive 96/9/EC on the legal protection of databases on the European level regulates rights in databases. This Directive aimed to harmonise part of copyright law dealing with databases and incentivise the development of the information society and investment in creating databases. Nevertheless, since this Directive was among the earliest ones in the copyright field, many issues arose over time with the development of the information society concerning the protection of the rights in databases. The development of open data policies is one of the most significant concepts that interfere with copyright and related rights, particularly with rights in databases. It will be discussed here whether the open data policies conflict with the protection of rights in databases or if they make convergence in which both concepts may exist and continue to develop together, side by side.

The idea of the protection of databases relies on two pillars: copyright and *sui generis* right in databases. Copyright in databases is an old concept which is upgrading the concept of the protection of collections. If collections are arranged in a systematic or methodical way and individually accessible by electronic or other means, they shall be considered databases. Collections may consist of independent works, data or other materials. They may be protected by copyright if they constitute the author's intellectual creation because of the selection or arrangement of their contents. So, databases will be protected by copyright if they are original and individual by selection or arrangement of their contents.

Nevertheless, originality and individuality will often be missing criteria related to databases because, usually, the principles of data organisation within the database will be simple and banal. Still, the content of the database will be much more interesting. For example, electronic databases may contain millions of items or data or information which are interesting to users because of data and not because of how data are organised within the database nor because only some data are selected among other possible data. So, data itself are taking much more attention than originality or individuality criteria in realising the database. Therefore, the EU developed a new concept in protecting databases by introducing so-called *sui generis* rights. This right does not protect the originality or individuality of a database creator but his investment. *Sui generis* right is turned towards the substantial investment of money, time, work or other valuable assets in data collection and database creation.

While the owner of the copyright database is entitled to exclusive rights to reproduction, distribution, communication to the public and adaptation concerning the whole database, the owner of the database protected by *sui generis* right is restricted to acts which include prevention of extraction and/or re-utilisation of the whole or a substantial part of the contents of a database. In the previous Copyright and Related Rights Act in Croatia, the protection of sui *generis rights* in databases was introduced in the law as a related right and included wide content of rights with respect to the entire content of a database, wider than envisaged by Directive 96/9/EC. On the contrary, in the new Copyright and Related Rights Act, the content of database rights regulated as related rights of a database producer in non-original databases is restricted to extraction and/or re-utilisation. This means that the content of rights in non-original databases is completely in line with Directive 96/9/EC and that the national legislator did not widen its content above this Directive. With this change, by narrowing the scope of exclusive rights in non-original databases, the legislator made a small step towards open data policies. Furthermore, this change concerning non-original databases opened more space for developing the freedom to operate. Namely, the concept of freedom to operate, in brief, gives any person the possibility to develop, make and market products without legal liabilities to third parties, such as owners of intellectual property rights. So, if the scope of the protection of non-original databases is narrower, it gives more space for developing and applying open data policies.

Moreover, in the new Copyright and Related Rights Act, a new concept is introduced, which regulates the status of copyright works and subject matters of related rights concerning employment contracts and contracts on commission. While previously it was presumed that the rights in copyright works and subject matters of related rights created under an employment contract or a contract on commission remains with the author and original owner of the related right, the new Act regulates a presumption that the employer and commissioner, respectively, are automatically the owners of the exploitation rights in works and other subject matters created under those contracts. This change should take away a part of the commercial risk from employers and commissioners who invest in creating copyright works and subject matters of related rights, support their better position in the digital market, and secure their portfolio of intellectual property rights towards third parties. But, of course, those presumptions are rebuttable. In addition, some other provisions are also regulated in the new Copyright and Related Rights, in the context of the mentioned two types of contracts, which should enhance a balance between the rights, interests and obligations of employers and commissioners, on the one hand, and the ones of authors and other creators, on the other. When applied to databases, this change should give database investors a better market position when they offer their databases in the digital market. But how will this change affect the open data policies?

On the one hand, the producers may better protect their databases by copyright and related right. On the other, they may, in a simpler way, turn to open data policies or use other possibilities to free their databases from copyright and related rights protection because their ownership is not questionable. The said refers, in particular, to public authorities, scientific and similar institutions, and private companies (in particular small and medium businesses) in Croatia and similar countries where the contracts on employment and commission in practical life, unfortunately, very often miss the specific provisions on intellectual property ownership.

Due to recent legislative changes in the Croatian Copyright and Related Rights Act, it may be reasonably expected that in the practical application, there will be much more situations where a related right of the producers of non-original databases will protect databases. On the other hand, databases will rarely be protected by copyright. In any of those two cases, the content of a database is not protected as such.

The described context of exclusive monopoly rights regulated to original and non-original databases shall apply to their authors and other creators, *i.e.* producers, employers and commissioners as derived owners of the same rights. The concept of exclusive and monopoly rights granted by copyright and related rights interferes with the interests of third parties. These interferences are even more intensive in a digital environment than in traditional circumstances. Namely, the business models, communication methods, and new approaches to data and privacy issues in the digital environment moved the boundaries of understanding the rights of privacy and private property, including intellectual property.

The idea of openness and free access to all information flooded the copyright and related rights. Open data policies have many faces. They vary from extreme viewpoints where everything that stays in the way of achieving absolute freedom of all information and absolute and divine right to free access and free use must be abolished, particularly intellectual property rights. More moderate views tend to find balance in the conflict between freedom of information and other rights and interests, such as intellectual property rights.

Open data policies rely on the constitutional right of freedom to information. The right to information is also regulated in the highest acts in the EU and the Council of Europe. In particular, open data policies rely on the idea that the information derived from public entities, institutions and other public authorities should be admissible without interference by public authority and regardless of the frontiers. Concerning this, the Croatian Right to Information Act mentions that its purpose is to enable and ensure the right to information as the constitutional right, as well as the right to re-use of information belonging to public authorities, including all public entities and trading companies. With this respect, it is to be seen how the intellectual property rights belonging to public authorities and other public entities and trading companies may survive and converge with the revival and renaissance of open data policies. It should not be forgotten that intellectual property rights are also constitutional rights.

The concept within the open data policies is the re-use of data collected, produced and developed by public authorities. This concept, in brief, relies on the idea that all data, all information created, produced, collected or analysed in the public sector, should be available for re-use for commercial or non-commercial purposes. This suggests that re-use should multiply the commercial value of the data and information created in the public sector for the direct or indirect benefit of society and its members. Furthermore, this sub-concept of open data policies relies on the idea that all public institutions should give up their intellectual property rights, particularly copyright and related rights, for the benefit of the concept of re-use. According to Directive 2019/1024 on the open data and re-use of the public sector data, the states should, by implementing this Directive, take a path towards the minorisation of copyright belonging to public authorities and other public entities, including public undertakings. For example, suppose there is copyright in data, *i.e.* information belonging to public sector bodies. In that case, the public sector body should diminish its

application and give the information for free or, if this is not possible, at marginal costs and only exceptionally at costs which would give a reasonable return on investment in the production, collection, or creation of the information. Moreover, it regulates that public sector bodies shall not exercise their *sui generis* database rights to prevent the re-use of documents or to restrict re-use beyond the limits set by that Directive.

It seems that the legislator on the EU level, when drafting the Directive 2019/1024, concluded that copyright belonging to the public bodies, particularly the database *sui generis* right, is in direct conflict with the open data policies and it should be narrowed, even put out of the application.

Also, Directive 790/2019 on copyright in the Digital Single Market gave another reason to rethink the protection of databases, particularly the non-original ones. Text and data mining exemption regulated in this Directive directly interferes with the copyright and related rights, particularly databases rights. This exemption directly affects the acts of extraction and re-utilisation, which is the basic content of rights in non-original databases. It regulates that database rights shall be restricted in favour of research organisations and cultural heritage institutions to carry out for scientific research, text and data mining of works and other subject matters to which they have lawful access. On the other hand, all other persons, including commercial undertakings, may benefit from the same exception if the right owner has not expressly reserved the rights concerning text and data mining. This exception to copyright and related rights is created to give free access to data for many purposes and the development of artificial intelligence, among other goals.

At the same time, legislators on the European and national levels take measures for better protection of copyright and related rights, particularly envisaged in Directive 790/2019. Taking into consideration that the same legislators at the same time are taking measures.

The shown examples conclude that the legislators on the European and national levels take measures to strengthen copyright and related rights in the Digital Single Market, particularly rights in databases, and promote and regulate open access concepts and principles. The idea seems to be to achieve convergence of those two concepts, despite their inherent conflict. New measures will probably be taken with the same aim. By now, the legislators have decided to favour open access in the public sector and apply the existence–exercise dichotomy. The copyright and related rights will not be abolished, and their content will not be squeezed.

Nevertheless, measures will be taken to make them not be exercised, sometimes by some soft and sometimes by more intensive legal tools. At this time, the public sector bodies are invited not to exercise their copyright and related rights. Furthermore, they are forbidden to exercise their *sui generis* or related rights in non-original databases for the benefit of open access and free re-use.

References

Blanke, H. J., Perlingeiro, R. (2018). Essentials of the Right of Access to Public Information: An Introduction. The Right of Access to Public Information An International Comparative Legal Survey, Springer, Berlin 33.

Charnley, P. (2012). European Court of Justice rules on the protection of databases under the Database Directive (96/9/EC). Computer Law & Security Review 28: 347-349.

Dalla Core, L. (2018). Towards Open Data Across the Pond. Open Data Exposed – Information Technology and Law Series, IT&LAW 30, Springer, 16, 22.

Derclaye, E. (2002). What is a database? The Journal of World Intellectual Property 5(6): 981-1011.

Derclaye, E. (2008). The Legal Protection of Databases A Comparative Analysis. Edward Elgar, Cheltenham, UK Northampton, MA, USA.

Derclaye, E. (2014). The Database Directive. In: (Stamatoudi, I., Torremans, P. eds.) EU Copyright Law – A Commentary; Edward Elgar, Cheltenham, UK • Northampton, MA, USA, 298-354.

Ducuing, C. (2020). Data as Infrastructure? A Study of Data Sharing Legal Regimes. Competition and Regulation in Network Industries 21(2): 6.

Elijaš, D., Marković, S., Trgovac, S. (2017). Pravo na pristup informacijama koje posjeduju tijela javne vlasti kroz ustavnosudski aspekt i stajališta Europskog suda za ljudska prava. Zbornik Pravnog fakulteta Sveučilišta u Rijeci 38(1): 501 – 526.

Forgo, N., Hinold, S., Schutze, B. (2017). The Principle of Purpose Limitation and Big Dana. New Technology, Big Data and the Law, Perspectives in Law, Business and Innovation, Springer Nature Singapore.

Galetta, D. U. (2018). Access to Administrative Documents and to Public Sector Information in Italy. The Right of Access to Public Information An International Comparative Legal Survey, Springer, Berlin, 360.

Giannopoulou, A. (2018). Understanding Open Data Regulation: An Analysis of the Licensing Landscape. Open Data Exposed – Information Technology and Law Series, IT&LAW 30, Springer, 106, 122.

de Hert, P., Sajfert, J. (2019). Regulating Big Data In and Out of the Data Protection Policy Field: Two Scenarios of Post-GDPR Law-Making and the Actor Perspective. European Data Protection Law Review 3(5): 338-351.

Klarić, P., Vedriš, M. (2014). Građansko pravo, Narodne novine.

Leistner, M. (2009). The protection of databases.In (Derclaye, E. ed.) The Research Handbook on the Future of EU Copyright, Edward Elgar, Cheltenham, UK • Northampton, MA, USA, 427-456.

Lassinatti, J. (2019). Re-use of public sector open data – characterising the phenomena. International Journal of Public Information Systems 13(1): 4. Available at: http://www.ijpis.net/index.php/IJPIS/article/view/158.

Leistner, M. (2018). Big Data and the EU Database Directive 96/9/EC: Current Law and Potential for Reform. In: (Lohsse, S., Schulze, R., Staudenmayer, D. eds.) Trading data in the digital economy: Legal concepts and tools ; Münster colloquia on EU law and the digital economy III. 1. Auflage. Baden-Baden: Nomos, 25-58.

van Loenen, B., Vancauwenberghe, G., Crompvoets, J., Dalla Corte, L. (2018). Open Data Exposed – Information Technology and Law Series, IT&LAW 30, Springer, 4, 16.

Marcowitz-Bitton, M. (2015). Commercializing Public Sector Information. Journal of the Patent and Trademark Office Society 97(3): 424.

Matanovac Vučković, R., Kanceljak, I. (2018). Pravo na ponovnu uporabu informacija javnog sektora prema autorskom pravu i srodnim pravima – sukob javnopravnog i privatnopravnog načela. Zbornik Pravnog fakulteta u Zagrebu 68(3-4): 372.

Meys, R. (2020). Data Mining Under the Directive on Copyright and Related Rights in the Digital Single Market: Are European Database Protection Rules Still Threatening the Development of Artificial Intelligence?" GRUR International 69(5): 11.

Musa, A. (2015). Od pristupa do ponovne uporabe informacija: Nove perspektive u kontekstu europeizacije. Zbornik radova: Znanstveno stručni skup s međunarodnim sudjelovanjem – Pravo na pristup informacijama i zaštita osobnih podataka, Split.

Musa, A. (2017). Pravni i institucionalni aspekti prava na pristup informacijama i transparentnosti u Republici Hrvatskoj. Spomenica profesoru dr. sc. Juri Šimoviću, Biblioteka suvremena javna uprava, Pravni fakultet Sveučilišta u Zagrebu, Studijski centar za javnu upravu i javne financije, Zagreb, 423 – 435.

Oesch, R. (2019). Der urheberrechtliche Schutz der Datenbanken und die Vertragsgestaltung ausfinnischer Sicht. Gewerblicher Rechtsschutz und Urheberrecht, Internationaler Teil - GRUR Int., 444-453.

Ofak, L. (2016). Pravo na pristup informacijama kao pravo zaštićeno Europskom konvencijom i drugim međunarodnim ugovorima za zaštitu ljudskih prava. Zbornik Pravnog fakulteta Sveučilišta u Rijeci 37(2): 921-951.

Otero, B. G. (2019). Evaluating the EC Private Data Sharing Principles - Setting a Mantra for Artificial Intelligence Nirvana? JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law 10(1): 68.

Rajko, A. (2010). Implikacije uvrštavanja prava na pristup informacijama javnog sektora u Ustav Republike Hrvatske. Hrvatska javna uprava 10(3): 629 – 649.

Richter, H. (2020). Exposing the Public Interest Dimension of the Digital Single Market: Public Undertakings as a Model for Regulating Data Sharing. Max Planck Institute for Innovation and Competition Research Paper 20-03: 10.

Salvadori, M. (2011). Right of Access to Documents: The Implementation of Article 42 of the Charter of Fundamental Rights. Frontiers in Artificial Intelligence and Applications, IOS Press, 39-53.

Spindler, G. (2019). Copyright Law and Artificial Intelligence. IIC - International Review of Intellectual Property and Competition Law 50: 1050.

Vanovermeire, V. (2000). The Concept of the Lawful User in the Database Directive. IIC international review of industrial property and copyright law 31(1): 63 – 81.

Wiebe, A. (2017). Schutz von Maschinendaten durch das sui-generis-Schutzrecht für Datenbanken, GRUR, 338-345.

Copyright and Related Rights Act NN 111/2019 (2021)

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (1996). Official Journal of the European Union, L 77.

Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information (2003). Official Journal of the European Union, L 345.

Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information (2013). Official Journal of the European Union, L 175.

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (2001). Official Journal of the European Union, L, 167.

Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (2019). Official Journal of the European Union, L, 130.

European Commission (2018). Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases; Available at https://op.europa.eu/en/publication-detail/-/publication/5e9c7a51-597c-11e8-ab41-01aa75ed71a1.

OPEN LEGISLATION: COMPARISON OF LEGAL PORTALS IN THREE COUNTRIES

Anamarija Musa[1], Margareta Habazin[1*], Tihomir Katulić[1]

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

*correspondence E-mail: mhabazin1@gmail.com

Keywords: Open Data; Legal data; Open legislation; Legislative portals

The digital transformation has affected people's lives, businesses, and the public sector. Large amounts of information and data from various sources have become available to the public (Jetzek et al., 2012). Nowadays, data plays a significant role in people's lives. Data has become the most critical driver of the digital age. Moreover, the public sector produces an enormous amount of data and information that are an extremely valuable resource and fuel that offers society an opportunity to drive tremendous economic and social change (Charalabidis et al., 2018). In spite of that, the opening of data by public organizations and public bodies has happened lately. Over the past few years, many governments have become driven and influenced by open data policies and have instructed their public organizations and public bodies to enable the release of public sector information and data (van Loenen et al., 2018). That has immensely increased the amount of open data freely available to the public. Although public organizations have made available their data, the potential of open data (data that anyone can access, share, use, and re-use without any restrictions or barriers) to become an essential tool for growth, innovation, and data-driven products and services can only be utilized when shared and re-used without restrictions or charge for their (re)use (Welle Donker and van Loenen, 2017).

Presently, vast collections of open data that hold enormous potential in various fields such as health, traffic etc. have been made available to the public without limitations for use and re-use (van Loenen et al., 2018). However, this research focuses on legal information and data and legislative portals as a specific type of open data portals that publish and make available legal information and data without restriction for their use and re-use. Legal information and data are specific types of open data that explain a law, legal system, and legal process and can be found in different legal documents such as acts, bills, case laws, administrative decisions, regulations, and similar (Musa, 2017). Policy-makers, legal experts and advisors, legal professionals, and legal scholars are constantly interested in legal data - one of the most crucial factors underpinning decision-making for various areas (Musa, 2017). In general, legal data is one of the essential sources of open data. Further, the availability of legal information and data is fundamental for legal certainty and legal and natural persons' conduct (Musa, 2017). The general public must be notified about laws and regulatory measures if they are affected by them, and legal documents with rules must be made publicly available to enable citizens and businesses to comply with them. Several thousands of legal acts and specific regulations are enacted by a national government and local authorities

every year. Also, on a daily basis, courts are issuing court rulings and decisions, and administrative bodies are making administrative decisions. Due to the digital age and fast-paced environment, the number of documents with legal data and information published online is growing exponentially (Musa, 2019). Thus, legislative portals as a specific type of open data portal are very important since they make legal information and data available to data users - legal professionals, lawyers, judges, legal researchers, legal scholars, and others. For legal professionals and other legal data users, it is essential to have the opportunity and convenience to easily and quickly find specific legal data and relevant legal documents that are fundamental for preparing a legal case, issuing the right decision, or just remaining informed about a specific legal topic. Also, law teaching and law student research resources have changed over the past 20 years and it is expected that many standard legal resources (legislation, case law, and other legal documents) are now available online.

Legal data holds enormous potential and promise, and legislative portals' functions and objectives should be the promotion and support of free access to legal data and information. Therefore, the goal of the research is to evaluate open data portals, governance, and availability of open data from the perspective of the legislative portal. All conducted research in the open data field acknowledges the relevance of appropriate governance structures, policies, and legal frameworks. Also, availability is one of the main requirements that need to be fulfilled to qualify data as open data. For that reason, the research tries to look into and determine the main features of governance of legislative portals in terms of key policy aspects, organization, and legal frameworks, and to what extent open data is available on the legislative portals. However, despite the significance of legal data as a specific type of open data, only a couple of scholars have paid attention to open legal data and legislative portals publishing these data. At the moment, no specific assessment instrument applicable to the open data legislative portals exists. In order to fulfill this goal, an assessment tool applicable to open data legislative portals based on the existing open data assessment theory and tools is created. The developed assessment tool has been applied to the legislative portals in specific countries chosen as a case study. The results of using the assessment tool show which similarities and differences between evaluated legislative portals exist.

Preliminary findings after applying the proposed assessment tool on selected legislative portals reveal that the effective development of legislative portals requires proper governance to organize and coordinate activities and contributions of different stakeholders. Also, findings prompt that legal data as a specific type of open data is considered available only if the data is a whole, downloadable from the Internet, with no costs, and can be freely (re)used or, in other words, if the data is made available in a format acceptable for further re-use following the open data standard. The initial analyses point out clearly that all assessed legislative portals aggregate collected legislative information from official sources and that the collected information is then standardized, cleaned, and published to the public. In addition, the analysis indicates that selected portals are very diverse in terms of performance and functionalities - each legislative portal has its own specific strengths and weakness. Besides, it is determined that companies that provide private legislative portals invest heavily in adding editorial content and developing sophisticated software for their search engines and interfaces. Thus, legislative platforms owned by private companies are unlikely to share the breadth of data necessary for robust scientific inquiry and public oversight. This research also identifies issues that prevent the effective re-use of legal data. Preliminary findings indicate that legislative portals should offer legal professionals and other legal data users a better quality of data

and make it easier for users to find reusable legal data and information. Also, access to legal data should be provided free of charge.

Consequently, this research states a set of related recommendations on how to overcome these issues related to the re-use of open data. Preliminary findings indicate that legislative portals should aim to serve legal professionals and other legal data users with better quality of data and the portal should strive to provide users with legal data in near-real time. Legislative portals should work as a catalyst triggering the publication of more and better quality legal data. Thus, in order that legal data can be re-used, more work needs to be done to offer all data in a machine-readable format. Because of the unique nature of legal content, special consideration must be given to the search engine, site navigation, and explanatory help. In addition, the availability of open data (especially technical openness and metadata completeness) should be improved and legislative portals should be established as web-based interfaces designed to make it easier to find reusable legal data and information. Further, legislative portals should support teaching and learning in legal education by creating and offering a database of important legal decisions freely and openly available to any interested legal data user. Consequently, the legislative portal's commitment to open access principles should provide that legal professionals, the business community, and the general public are able to access the most important legal documents, data, and materials free of charge.

Acknowledgements: The authors acknowledge the financial support from TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, Th.J., Ferro, E. (2019). The World of Open Data: Concepts, Methods, Tools and Experiences. Springer Cham.

Jetzek, T., Avital, M., Bjørn-Andersen, N. (2012). The Value of Open Government Data: A Strategic Analysis Framework. Paper presented at 2012 Pre-ICIS Workshop, Orlando, Florida, United States.

Musa, A. (2019). Open data and re-use of public sector information: What does the Directive on Open Data bring to citizens and public sector? Informator: instruktivno-informativni list za ekonomska i pravna pitanja 6604: 2-12.

Musa, A. (2017). Transparency of judiciary in the function of the efficient law application and effective public administration: the publication of the decisions of administrative courts. Hrvatska I komparativna javna uprava/Croatian and Comparative Public Administration 17(4): 3-45.

van Loenen, B., Vancauwenberghe, G., Crompvoets, J., Dalla Corte, L. (2018). Open data exposed. In van Loenen, B., Vancauwenberghe, G. and Crompvoets, J., eds., Open data exposed. T.M.C. Asser Press, The Hague.

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

SESSION VI

GEOSPATIAL REASONING WITH OPEN DATA

ASSESSING AN OPEN SPATIAL DATA INFRASTRUCTURE FROM A USER PARTICIPATION PERSPECTIVE: A QUALITATIVE EXPLORATORY RESEARCH WITH OPENSTREETMAP

Guilherme Spinoza Andreo[1*], Frederika Welle Donker[1], Stefano Calzati[1]

[1] Faculty of Architecture & the Built Environment, Delft University of Technology, Netherlands

*correspondence E-mail: guispinoza@gmail.com

Keywords: Open Spatial Data Infrastructure (SDI); OpenStreetMap; User participation; Volunteered Geographic Information (VGI); Motivational Factors

1. Introduction to Open Spatial Data Infrastructures

1.1. Spatial Data Infrastructure developments

Governments have invested in establishing Spatial Data Infrastructures (SDIs) to facilitate sharing spatial data between government organisations. An SDI is a framework of policies, access networks, data, standards and individuals that promote, coordinate and facilitate the dissemination and use of open spatial data for users in the spatial data community (Rajabifard and Williamson 2002). Initially, production and distribution of spatial data was limited to national mapping authorities, with access to SDIs limited to public sector bodies. Soon after, the way in which spatial information was used, produced and shared changed dramatically. The recent evolution and advances in Global Navigation Satellite Systems (GNSS), mobile devices and Global Positioning System (GPS) enabled devices allowed for common citizens to participate in the GI experience (Gómez Barrón Sierra, 2020; Costa Fonte et al., 2017; Budhathoki, 2010), which also became known as Volunteered Geographic Information (VGI) (Goodchild, 2007), such as is the case with the OpenStreetMap (OSM) project.

1.2. Open Spatial Data Infrastructure Ecosystems

Similarly, a new trend for SDIs have emerged as a more 'open' SDI. In an open SDI, citizens, businesses, academics and non-governmental actors can share their data as stakeholders of the infrastructure (Vancauwenberghe et al., 2018; Vancauwenberghe and van Loenen, 2018), while also providing an increased availability for participating in important governance and organisational decisions.

Recently, the public sector has taken the initiative to adopt open data formats to ensure that data are provided for free, are machine-readable, and have no licence restrictions (Vancauwenberghe and van Loenen, 2018). Open formats facilitate the connection of GI with different providers that use the same standards. This enables people to innovate, use or create applications and Application

Programming Interface (API) as they please and to engage in public decisions (Fox, 2013; Varga et al., 2022).

1.3. How can OpenStreetMap contribute to Open SDI research

Citizens can supply valuable VGI to SDIs at a high technical level (Goodchild, 2007), as is the case with OpenStreetMap (OSM). This can lead to improved levels of participation of users in SDIs (Mooney and Corcoran, 2011). The quality of VGI and its integrity has been researched previously (e.g., Mashhadi et al., 2015), as well frameworks to increase contribution to the public sector (e.g., Khan and Johnson, 2020), and the factors to contribute to OSM. However, research into which factors contribute to ensure long-term user participation in Open SDI has yet to be done. We need to understand the status of spatial data in open spatial data infrastructures and to identify the factors that motivate users and how user participation in an open SDI can be assessed properly.

In this article, we answer our research question '*What are the factors that foster user participation in the community of an open SDI*?' Our exploratory research shows how theory can be put into practice to move open SDIs to the next level, as well as to understand how the extrinsic, and intrinsic motivational factors can build, or otherwise hinder, a sustainable participatory community. To do so, we use OSM as a case study to assess key success factors that can be used to foster user participation in open SDIs. We consider the intrinsic and extrinsic factor that motivate user participation in OSM, how and to which extent our selected users participate, and what can be done to motivate the community of users to participate in an open SDI.

2. Literature review into user participation factors

User participation indicates how much people are willing to engage on a personal and organizational level to contribute their knowledge on specific issues. User participation in relation to spatial data contributions vary according to the amount of social pressure to be involved or to have a sense of inclusion in important components within organizations (Montalvo, 2003, cited by Rajabifard et al., 2006). This can result in a limited engagement of the actors that are responsible for promoting data reuse in distinguished initiatives. Other studies have shown that some of the factors that play an important role in open data user participation are mainly the quality of infrastructure. the knowledge shared, the confidence of users in the open data, how useful it was and if it met their expectations (Krismawati and Hidayanto, 2021).

Although OSM is not by any means the sole repository for VGI data, it is one of the most common repositories. OSM was initially created in 2004 to map streets to overcome licensing restrictions of certain maps at the time and to supply crowdsourced geographical information for users (Bennett, 2010). The purpose for which it was made was intrinsically connected to gathering freely provide available spatial data, which was previously restricted for small businesses, individual users and community organizations that could not afford, access or modify the traditional GI that was provided through the data steward (Budhathoki and Haythornthwaite, 2013).

Motivations to contribute to VGI is influenced by the technology that people have available, such as GPS enabled devices, computers storage and processing capacity enable people to feel motivated to voluntarily give their collected data (Tanaka, 2017). The most important component in VGI is the collective effort of the contributors, which differs from the traditional effort to produce spatial

data, since there is no monetary compensation or someone to direct you on how to contribute. This way of producing spatial data resembles a Wikipedia style for generating new information (Budhathoki et al., 2008; Budhathoki and Nedović-Budić, 2010; Tanaka, 2017). According to Nielsen (2006), within online communities and social media, only 1% of people actively contribute, 9% contributes occasionally and 90% of users do not contribute at all and are identified as 'lurkers'. Nielsen (2006) defines this as "participation inequality", and within Wikipedia, this inequality is even more disparaging with 99,8% of the population characterized as 'lurkers' (Sjoukema, 2015).

3. Research methodology

This research used preliminary web-based surveys and semi-structured interviews with users of OpenStreetMap to explore the factors that foster user participation in the community of an open SDI. We used a methodology that adapted motivational factors from Budhathoki (2010). In relation to the motivational factors, intrinsic motivation is derived from inner core values that motivate someone to participate, since it is not reliant on external factors from the outside world, whereas extrinsic motivation is related to tangible external factors that surround us, such as financial rewards or outside recognition (Hennig, 2020). This formed the theoretical framework that was used to development of the qualitative assessment of the user participation for Open SDI. After joining and establishing contact with the community of OpenStreetMap, preliminary surveys were distributed within the active user communities, which were used to contact volunteers for further semi-structured in-depth interviews. Then, a qualitative analysis of the surveys and the interviews was executed to examine what are the key themes, indicators and motivations that were identified from the users to participate in an Open SDI. A total of 100 valid responses from the survey were recorded and 13 interviews were conducted with survey participants who agreed to be interviewed, plus two interviews with academic experts in SDI user participation.

4. Analysis of our research

Our results showed that most survey respondents participate in OpenStreetMap on average at least a few times a week or every day. This highlights that the volunteers that participated in the survey are most likely "heavy contributors" to the project. The main motivations identified in the survey (response rate), were extrinsic factors of believing in OSM's project goal, OSM community and career, and the intrinsic factors of fun, instrumentality of knowledge, and altruism. It is important to mention that users felt very strong about the positive impact the community has, not only in the project, but also as a feedback loop to the open data ecosystem which OSM is a part of. Simultaneously, it was also the main point of improvement for OSM according to some participants. Most of what participants recommended were generalized improvements for new users, like what Budhathoki and Haythornthwaite (2013) describe lightweight organization and collaboration to be. However, it shows that heavy contributors have also lightweight contributor motivations, which differs from what has been shown in the previous studies into motivational factors of OSM.

Mixed feelings were shown in relation to the concern of commercial and institutional influences inappropriately taking over OSM (over the policies, standards and restrictive licensing for external use). This is in alignment with the anti-corporate sentiment presented by Budhathoki and Haythornthwaite (2013) and Budhathoki and Nedović-Budić (2010). Users generally trust the

infrastructure, even though there could be potential improvements and especially the "heavy contributors" feel the trust, design and self-organization in the project, despite the interference of some users who do not follow the guidelines. In general, users understand the benefits of open data within the OSM ecosystem but believe that establishing a harder standard for the OSM data would improve restrictions and licensing barriers for external users (both for the public/private sector and academia). Respondents recognize the role of OSM in the larger open data ecosystem (i.e., public sector, countries, National SDIs) and the open data ecosystem that envelopes OSM data. Both experts and interviewees acknowledge the underlying issues with OSM's licenses and standards, although linked data should be possible, if a direct communication line exists with external stakeholders.

As to how and to what extent OSM users participate in the open data community, most of the survey respondents were users that contribute at least a few times a month and on average a few times a week. This could be either due to users in OSM participating as much, or most likely due to the volunteer bias. Since many users also showed high interest to learn, participants in this research did so because they believed the research addresses important challenges, such as how to maintain motivation in the long-term. Nevertheless, further research could shed light on this matter.

Finally, factors that could foster user participation in an open SDI in general, as proposed by OSM users, are better support mechanisms, tutorials for the multiple editors of OSM as an open SDI and centralization of one main communication channel as a reliable source of information. Moreover, more accessible editing on digital devices would also enhance user participation since users could contribute from more devices. Better user guidelines would also help, not only new users but also already contributing users. OSM is quite a complex platform with many factors that must be considered before adding new data and, therefore, there are guidelines and standards suggested by the community for how to do so in a uniform and accepted way. Currently, users mention that the lack of prescriptive guidelines for adding new data provides such a steep learning curve that not all participants are able to to understand. Hence, a less overwhelming step could be included to bridge this gap either with more tutorials or videos.

5. Conclusion and recommendations

In summary, the belief in the project goal of OpenStreetMap, the perceived importance in the community by the users to the development of OpenStreetMap and its ecosystem, fun, instrumentality of knowledge, altruism, unique ethos, meeting their self-needs and learning are among the most important factors for active users to participate in an Open SDI. To engage both potential new and long-time users in an efficient, user friendly, and a light-weight manner requires balancing priorities in relation to the infrastructure's current status and goals. Maintaining an active community requires engagement on a personal level, where transparent or more enforcing user guidelines would also help not only new users, but also the ones that already contribute. OpenStreetMap is perceived as an Open Data Ecosystem among its users and as an Open SDI amongst the selected experts, yet there were dividing opinions to the perception about Open Spatial Data Infrastructures in relation to OpenStreetMap from their own users. Essentially, by itself, OpenStreetMap provides only map data, which would be another contention point for determining if it is an open SDI, although vector tiles are in development in the OpenStreetMap infrastructure.

Users are free to participate openly in many different fields of OpenStreetMap if desired, i.e., any of the five key components of SDI: standards, policies, access network, data, and governance.

Acknowledgements: The authors acknowledge the financial support from TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Bennett, J. (2010). OpenStreetMap. Packt Publishing Ltd. Available at https://www.perlego.com/book/389235/openstreetmap-pdf

Budhathoki, N. R. (2010). Participants' motivations to contribute geographic information in an online community. University of Illinois at Urbana-Champaign.

Budhathoki, N. R., Haythornthwaite, C. (2013). Motivation for open collaboration: Crowd and community models and the case of openstreetmap. American Behavioral Scientist 57(5): 548-575. https://doi.org/10.1177/0002764212469364

Budhathoki, N. R., Nedović-Budić, Z. (2010). An interdisciplinary frame for understanding volunteered geographic information. Geomatica 64(1): 11–26. https://doi.org/10.5623/geomat-2010-0003

Costa Fonte, C., Fritz, S., Olteanu-Raimond, A.-M., Antoniou, V., Foody, G., Mooney, P., See, L. (2017). Mapping and the citizen sensor. Ubiquity Press.

Goodchild, M. F. (2007). Citizens as voluntary sensors: spatial data infrastructure in the world of web 2.0. International journal of spatial data infrastructures research 2(2): 24–32.

Gómez Barrón Sierra, J. P. (2020). A Socio-Technical Design Approach to Build Crowdsourced and Volunteered Geographic Information Systems: Leveraging the Crowds and Participatory Communities for Geoinformation Management. Doctoral dissertation.

Hennig, S. (2020). Motivation and its consideration in participatory spatial data contribution. The Professional Geographer 72(2): 238–252. https://doi.org/10.1080/00330124.2019.1676799

Khan, Z. T., Johnson, P. A. (2020). Citizen and government co-production of data: Analyzing the challenges to government adoption of VGI. The Canadian Geographer/Le Geographe canadien 64(3): 374–387. https://doi.org/10.1111/cag.12619

Mashhadi, A., Quattrone, G., Capra, L. (2015). The impact of society on volunteered geographic information: The case of OpenStreetMap. In OpenStreetMap in GIScience 125–141. Springer.

Mooney, P., Corcoran, P. (2011). Can volunteered geographic information be a participant in environment and SDI? In International Symposium on Environmental Software Systems 115–122. Springer.

Nielsen, J. (2006). The 90-9-1 rule for participation inequality in social media and online communities.

Rajabifard, A., Williamson, I. P. (2002). Spatial Data Infrastructures: an initiative to facilitate spatial data sharing. Global environmental databases-present situation; future directions, 2.

Sjoukema, J. (2015). Keeping the BGT up-to-date: Volunteered geographic information, a serious option? Master thesis. Utrecht University.

Tanaka, Y. (2017). Motivation for Contribution to Volunteered Geographic Information in Nepal. Master thesis. University of California, Davis.

Vancauwenberghe, G., Valečkaitė, K., van Loenen, B., Welle Donker, F. (2018). Assessing the openness of spatial data infrastructures (SDI): Towards a map of open SDI. International Journal of Spatial Data Infrastructures Research 13: 88–100.

Vancauwenberghe, G. van Loenen, B. (2018). Exploring the emergence of open spatial data infrastructures: analysis of recent developments and trends in Europe. User Centric E-Government 23–45.

Varga, F., Hrustek, L., Kević, K., Welle Donker, F., Šalamon, D. (2022). Urban dog spaces: The openness of dog-related government data in the city of Zagreb, Croatia. Interdisciplinary Description of Complex Systems 20(2): 125–135. http://dx.doi.org/10.7906/indecs.20.2.5

OPEN SPATIAL DATA INFRASTRUCTURE ACTIVE LEARNING AND TEACHING METHODS IN PRACTICE

Frederika Welle Donker[1*], Bastiaan van Loenen[1], Vesna Poslončec Petrić[2]

[1] Faculty of Architecture & the Built Environment, Delft University of Technology, Netherlands
[2] Faculty of Geodesy, University of Zagreb, Croatia

*correspondence E-mail: f.m.welledonker@tudelft.nl

Keywords: spatial data infrastructures; Open SDI; education; active learning and teaching methods

1. Introduction to open spatial data infrastructure education

1.1. Open Spatial Data Infrastructures

There is an increasing need for spatial data to be used for informed decision-making and as a resource for developing innovative products and services. A Spatial Data Infrastructure (SDI) facilitates access to and sharing of spatial data by providing a framework in which technical and non-technical aspects are established. Traditionally, SDIs are developed by and for government organisations to share spatial information within government. Non-government parties may also use SDIs to access government spatial data, either as open data or as fee-based data. There are also non-government SDIs, such as company SDIs used to share spatial information within a private sector organisation. OpenStreetMap is another example of a non-government SDI, whereby citizens have created a worldwide map based on crowd-sourced information and contributions from private and public sector organisations. The concept of Open SDIs has emerged from an increased interest in open data initiatives steered by national and international directives, such as the EU Open Data Directive (Directive (EU) 2019/1024), as well as the large investment of European public authorities in developing SDIs. Open SDIs are SDIs in which citizens, research institutions, private organizations and other businesses and non-governmental actors are recognized as key stakeholders of the infrastructure (Vancauwenberghe and van Loenen, 2018; Vancauwenberghe et al., 2018). The concept of Open SDIs is about openness to new stakeholders in the open data ecosystem, in which stakeholders can be both producers and users of spatial data. Open SDIs are also linked to developments and trends in other domains, such as open government, open data, open science, and open software. This new paradigm regarding Open SDIs means that new particular skills are required, which currently are not offered by traditional SDI education. Open SDI education requires a shift from teaching skills in single disciplines to a multi-disciplinary approach.

In addition to the content of the SDI education, a need towards more active teaching and learning in SDI arose. Several universities introduced this type of education for single topic classes but until

recently, there was barely any exchange of concepts and experiences, let alone guidelines or best practices on active teaching and learning on SDI.

1.2. The SPIDER Project

The Erasmus+ open SPatial data Infrastructure eDucation nEtwoRk (SPIDER) project integrates both developments by promoting and strengthening active learning and teaching towards Open SDIs. This paper describes how we researched, developed and implemented the concept of Open SDI as a new paradigm to SDI education. The SPIDER partners – Bochum University of Applied Science, Delft University of Technology, KU Leuven, Lund University, and the University of Zagreb – explored which active learning and teaching methods are already implemented in their SDI courses, and which methods can still be implemented.

1.3. Reading Guide

In Section 2, we provide an overview of our research of active teaching and learning methods. Section 3 describes the methodology we used to adapt and implement active learning and teaching methods to an online environment. We also describe selection processes for developing new topics to be implemented in SDI teaching. In Section 4, we show how we tested the newly developed active learning and teaching methods during the SPIDER Summer School, held 22-26 August 2022 in Zagreb. Section 5 provides our conclusions, reflection, and recommendations for active teaching in SDI courses.

2. Active Teaching and Learning Methods

Research has shown that the effectiveness of passive teaching methods may be questioned (Bonwell and Eison, 1991; Renkl et al., 2002; Michel et al., 2009). Research has also shown that students participating in active learning activities are engaged in higher-order thinking tasks such as analysis, synthesis and evaluation, and learn more than they do compared to traditional lectures (Deslauriers et al., 2019). Active Learning and Teaching is a broad concept, which refers to methods that will dynamically involve students in the learning process (Menekse et al., 2013). There are many different methods to actively engage students in the learning process, ranging from very short, e.g., one-minute papers, to long, e.g., problem based learning, and ranging from using traditional media to using multimedia technologies. Active learning activities can be especially effective in engaging students at higher learning levels according to Bloom's Taxonomy, as these higher levels are less attended in teaching methods (Welle Donker et al., 2022). According to this taxonomy, students engage lower levels of learning by remembering and reproducing information. At higher levels of learning, students are able to apply this knowledge and to analyse. At the highest levels of learning, students are able to critically evaluate information and to create new work, see https://www.bloomstaxonomy.net for more details. It should be noted that active learning is not limited to the classroom, but can also be performed outside the classroom. **Table 1** shows an overview of active learning activities categorised by learning levels according to Bloom's Taxonomy and for the different teaching environments.

Table 1. Active learning activities categorised by learning levels according to Bloom's Taxonomy and by teaching environment (source: Welle Donker et al., 2022)

Learning level according to Bloom's Taxonomy	on-campus teaching session	online teaching session	outside teaching sessions
Remember	demonstrations examples guest speakers in-class quizzes/polls	demonstrations examples guest speakers in-class quizzes/polls	clips podcasts class recordings short quizzes/ self-tests
Understand	asking questions active listening / paraphrasing one-minute paper / one-sentence summary brainstorm / brainwrite jigsaw in-class quizzes/polls mind map	asking questions active listening / paraphrasing one-minute paper / one-sentence summary brainstorm / brainwrite jigsaw in-class quizzes/polls mind map	literature / reader short quizzes / self-tests mind map
Apply	debate student presentations concept map	debate student presentations concept map	exercises serious games concept map
Analyse	muddiest point concept map / mini map active writing class discussions cases / role play / simulation think-pair-share / turn & talk / snowball group investigation as collaborative learning	muddiest point concept map / mini map active writing class discussions cases / simulation think-pair-share group investigation as collaborative learning	concept map
Evaluate	peer instruction peer review peer tutoring classroom quizzes	peer instruction peer review peer tutoring classroom quizzes	self-tests portfolio
Create	formulating exam questions mini lectures	formulating exam questions mini lectures	research paper case study / project formulating exam questions

3. Methodology

As part of the SPIDER Project, we carried out a literature review into active learning and teaching methodologies and evaluations in 2020-2021. In addition, we made an inventory of which methodologies were already implemented in SDI teaching at the SPIDER partners' universities. We carried out an evaluation of these methodologies in 2021 to assess in which way these practices were adapted to teaching in an online environment during the COVID-19 lock-downs. In our

evaluation, we included both students' assessment as well as teachers' assessment. In parallel, we organised a workshop for SDI education stakeholders in January 2022 in which we collected an overview of topics that should be included in Open SDI teaching. **Figure 1** shows an overview of the suggested topic and their relevance. The lessons learned were included in the development of new active learning and teaching methods, which were tested during the SPIDER Open SDI Summer School 22-26 August 2022 in Zagreb.

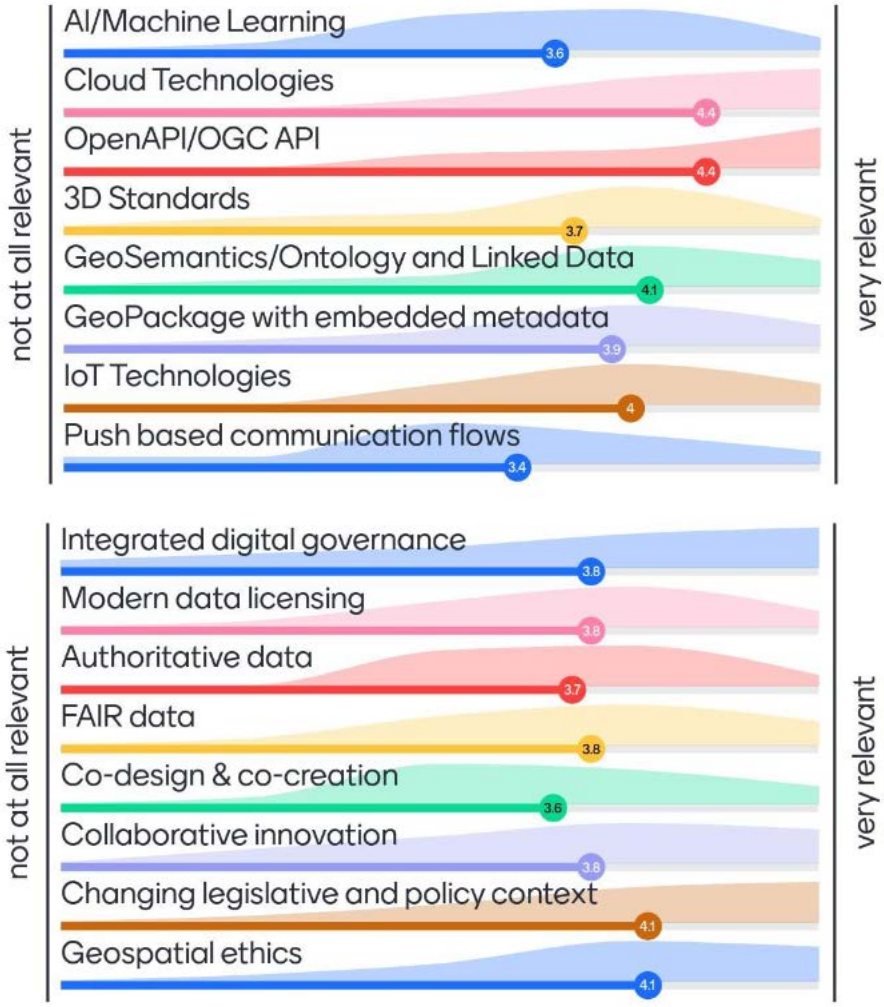

Figure 1. Suggested topics to be included in Open SDI teaching (source: Vancauwenberghe et al., in press).

A number of these topics, such as FAIR data and modern data licencing, are already covered in SDI courses of some of the SPIDER partners. Therefore, the SPIDER partners developed 10 topics that are new to the SPIDER partners' SDI courses. We tested the active learning and teaching methods that we intend to employ during the SPIDER Summer School 22-26 August 2022 in Zagreb. In this paper, we describe three of these topics: Private Sector Data in the Open SDI; Crowd-sourcing Spatial Data & Citizens' Science; and Ethical issues in Open SDIs.

4. Newly developed topics and active learning methods

4.1. Private Sector Data

In traditional SDIs, the private sector's role was primarily a user rather than a provider. With the economy becoming more data-driven, many companies and non-government organisations are creating their own SDI, e.g., Google, and OpenStreetMap. Some companies already made their spatial data or infrastructure openly available, e.g., Fugro contributing bathymetry data or Google making their infrastructure openly available during disasters to allow real-time updates. The European Union envisions a major role for G2B and B2G mechanisms and has introduced the concept of common data spaces that should ensure that more data becomes available for use, "while keeping the companies and individuals who generate the data in control". During the Summer School, students engaged in role plays, in which students represented different stakeholder groups. In one role play, students were provided with a description of the characteristics (e.g., responsibilities, funding, attitudes, desires) of "their" stakeholder (government policy makers, government data providers, European Commission, private sector SMEs, private sector big tech, civil society, and academia). In another role play, the stakeholders had to argue whether they should be included in the development of Open SDIs, or whether non-government organisations should be compelled to contribute their data to an Open SDI.

4.2. Crowdsourcing GI and Citizens Science

Nowadays, everybody can contribute to collecting data to monitor societal problems, such as noise pollution or air quality. Required technology becomes more accessible as smartphone apps and their hardware becomes more powerful to be used to gather data. Examples are GPS, sound meters and smartphone adaptors to measure air quality. The collected data can be a valuable contribution to Open SDIs. During the Summer School, groups of students went out on the street to collect noise data with a mobile app. The data, collected in different formats, were imported into GIS software, classified, and manipulated to visualise the data in a map, including elements such as coordination grid and various annotations. Not all students had any previous experience with data acquisition or data manipulation. These activities stimulated peer-assisted data acquisition and data manipulation as students had to decide as a team what should be included in the map.

4.3. Ethical Challenges of Open SDIs

With more data being collected and used for decision-making for society, there is a risk of unintended effects on individuals. More data or technology does not necessarily lead to better solutions; the context of "better" is extremely contextual, e.g., policing algorithms gone wrong. To demonstrate these complex issues, the Summer School students were introduced to the so-called trolley problem: a train trolley is moving on the tracks. The brakes of the trolley are broken. If the students do nothing the train will continue on the current tracks. If, however, students pull the lever, the train will be directed to a different track. The typical scenario is that not pulling the lever will kill five people, while pulling the lever will only kill one. Students were divided into distinct groups and had to decide whether to pull the lever given different scenarios. This resulted in heated debates on who should be spared and who should be sacrificed. As a second exercise, the students had to apply the Ethics Assessment List for Geoinformation Initiatives (EALGI) and test their case studies. EALGI was not specifically designed to be applied to open SDI, but retrospectively

it proved flexible enough to fit the task. Students were given half an hour for delivering a comprehensive application of EALGI to their case study, then provided feedback on the robustness of EALGI, and finally presented their results.

5. Conclusions, reflections, and recommendations

The SPIDER Summer School provided a good platform to test active learning and teaching methods for newly developed topics as the students were from diverse backgrounds. Although we only tested these activities in a classroom environment, parts of these activities can be adapted to an online environment or as activities outside the classroom. The crowdsourcing data acquisition proved to be very suitable as a team building exercise, important for future group work. The active learning activities also showed that activating the students is not only especially useful for transferring knowledge, but also a lot of fun.

Acknowledgements: This paper was written as part of the SPIDER: open SPatial data Infrastructure eDucation nEtwoRk project under the Erasmus+ programme (project number 2019-1-DE01-KA203-005042). The authors also acknowledge the financial support from TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Bonwell, C. C., Eison, J. A. (1991). Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports [monograph]. ERIC Clearinghouse on Higher Education, The George Washington University. Available at https://eric.ed.gov/?id=ED336049

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. Proceedings of the National Academy of Sciences 116(39): 19251-19257. https://doi.org/10.1073/pnas.1821936116

Menekse, M., Stump, G. S., Krause, S., Chi, M. T. H. (2013). Differentiated Overt Learning Activities for Effective Instruction in Engineering Classrooms. Journal of Engineering Education 102(3): 346-374. https://doi.org/https://doi.org/10.1002/jee.20021

Michel, N., Cater III, J. J., Varela, O. (2009). Active versus passive teaching styles: An empirical study of student learning outcomes. Human Resource Development Quarterly 20(4): 397-418. https://doi.org/https://doi.org/10.1002/hrdq.20025

Renkl, A., Atkinson, R. K., Maier, U. H., Staley, R. (2002). From Example Study to Problem Solving: Smooth Transitions Help Learning. The Journal of Experimental Education 70(4): 293-315. https://doi.org/10.1080/00220970209599510

Vancauwenberghe, G., Lacroix, L., Casiano Flores, C., Welle Donker, F., van Loenen, B., Calzati, S., Keßler, C., Küppers, N., Panek, M., Mansourian, A., Zhao, P., Tomić, H., Kević, K., Poslončec Petrić, V., Bačić, Ž. (in press). Development of new active teaching practices SPIDER Project.

Vancauwenberghe, G., Valečkaitė, K., van Loenen, B., Welle Donker, F. (2018). Assessing the Openness of Spatial Data Infrastructures (SDI): Towards a Map of Open SDI. International Journal of Spatial Data Infrastructures Research 13: 88-100. https://doi.org/10.2902/1725-0463.2018.13.art9

Vancauwenberghe, G., van Loenen, B. (2018). Exploring the Emergence of Open Spatial Data Infrastructures: Analysis of Recent Developments and Trends in Europe. In S. Saeed et al. (Eds.), User Centric E-Government 25. Springer International Publishing AG. https://doi.org/10.1007/978-3-319-59442-2_2

Welle Donker, F., van Loenen, B., Keßler, C., Küppers, N., Panek, M., Mansourian, A., Zhao, P., Vancauwenberghe, G., Tomić, H., Kević, K. (2022). Showcase of Active Learning and Teaching Practices in Spatial Data Infrastructure (SDI) Education. AGILE: GIScience Series 3: 18. https://doi.org/10.5194/agile-giss-3-18-2022

ENGAGING WITH OPEN DATA THROUGH VISUALISATION AND COMMUNICATION: THE ROLE AND POSSIBILITIES OF CARTOGRAPHY

Ana Kuveždić Divjak[1*], Karlo Kević[1]

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

*correspondence E-mail: akuvezdic@geof.hr

Keywords: open data; mapping visualization; map; cartography; engagement

Many uses of open data rely on being mapped – that is, provided with a spatial reference to their particular location on Earth. For example, various socio-economic indicators can be mapped on top of residential districts. Soil quality data can be linked to digital elevation models to model erosion potential, and the same soil data can be compared to geographically overlapping land ownership and land subsidy data. Because human activities are largely confined to the near-surface of the Earth, many open datasets would have far less impact without their spatial component.

By applying different visualisation techniques and tools, various socio-environmental problems can be analysed more deeply, because many sets of data can be observed together, regardless of whether they refer to the physical or social world. For example, to illustrate the effect of a green infrastructure development program, instead of images and tables "locked" in a PDF document, the visualisation of this data on a map can provide a common basis for experts from different domains in researching public policy approaches to mitigate the effects of heat islands in urban environments.

In an era of open data, visualisation tools and technologies are essential to analyse the large amount of available data sets and to make data-driven decisions. Effective visualisation can help in the effort to make mutual relationships obvious, to recognize the impacts of different actions more clearly, and to model and evaluate compromises between possible solutions. As a result of the visualized data being comprehensible to a wide audience, public policy makers can use the visualisations to engage and use the collective intelligence of different stakeholders in the collaborative and democratic development of solutions (Nash et al., 2022).

Although the concept of data visualisation is not new, developments in technologies and analytical tools have changed the way visualisation is done and expanded its application and relevance. With the widespread use of data and statistics for planning and commerce in the 19th century, the practice of transforming quantitative data into visual results has its roots in cartography. Ptolemy's map, known in the 2nd century, marked geographic longitude and latitude with lines, visualizing for the first time phenomena that did not physically exist.

Today, with open data and open technologies, the production of mapping visualisations is becoming more accessible than ever. Although graphic designers and cartographers still play an important role, data-driven communications are not the exclusive domain of these professionals. Instead, they are increasingly used by a variety of knowledge workers (including developers, researchers, journalists, civil society organisations, and businesses) – for their own use or to inform others. As a result, mapping visualisations have been identified as a strategic advantage for the popularity of open data, and they continue to be one of the entry points for the public to engage with open data.

Although the specific goals may vary (e.g., depending on the specific societal question being addressed), the overarching goal of these maps remains the same: to make complex issues understandable to a broad audience and to persuade users of the mapping visualisations to engage and take action.

However, the creation and use of maps for such efforts is not straightforward. Research (e.g., Chess and Johnson, 2007; Ansari et al., 2022; Fish, 2020; Kraak, 1998) has shown that simply presenting facts is often insufficient for someone to take action to address these types of socioecological problems. Not only must the information be presented accurately, but it must also appeal to readers' emotions (Fish, 2021). Design defaults or visualisation limitations in common mapping software packages and application programming interfaces (API) can result in unintended communication consequences (Ricker, 2020).

Therefore, it is important to examine the conditions under which such mapping visualisations are created and disseminated and are thought to foster processes of meaning making, learning, and engagement. Understanding who creates these mapping visualisations and what techniques they use to make them engaging or not is important to assessing the power of visualisation to multiply the impact of open data.

We explore the following questions: Who produces and disseminates mapping visualisations based on open data? Were they created in-house or reproduced from other sources? What themes/topics and geographic extents are represented in the mapping visualisations? What types of cartographic methods (e.g., map type and visual variables) were used, and how successful were the methods used in moving beyond mere description and reporting to engaging and inspiring action. Answering these questions will shed light on current trends in open data visualisation and better illustrate the ways in which modern cartography can better engage and serve citizens through visualisation and communication.

The research methodology included the identification of use cases for the visualisation of open (government) data and the analysis of their characteristics. The basic criterion for selecting use cases was that the visualisation is openly available on the Internet and created with open data (as the main source). The use cases for mapping visualisations were found using search engines under the keywords "Open Data Mapping Visualisation", "Open Data Map", "Open Data Cartographic Visualisation", "Open Government Data Map" and "Open Government Data Visualisation". Based on the requirements, 23 use cases for open data visualisations were found at different organizations. These included public organizations at different levels of government (EU, national, regional, local), businesses, research institutions, and non-profit organizations.

The evaluation of the collected maps is based on content analysis, a systematic approach already used in cartographic research to analyse and compare maps (Muehlenhaus, 2011; Fish, 2017; Roth, 2015). The content analysis method is based on the identification of a set of codes that serve as operational rules that define the definitions and intensity of the different components in the maps analysed. Ideally, the use of codes ensures that each map is systematically evaluated and analysed in exactly the same way so that the results can be compared. It also allows for further analysis by other researchers in the future.

The coding scheme was created to understand the content contained in the maps in terms of: the source of the data, the target group for which it is intended for, who provided the data for use, under what conditions (license) the data are available, and the characteristics of the cartographic visualisation: static or dynamic - interactive or view only, structure type (sequential or non-sequential), and type of navigation: thematic, spatial, and temporal. It was also investigated whether accessibility for the visually impaired or the mechanisms for sharing and embedding content in different environments were implemented.

Preliminary results show that a total of 21 maps contain open government data, while only two maps are based on data collected data. Since maps can have multiple target groups, 17 use cases are for citizen use, 13 are for expert use, and 12 are for government organizations and institutions. As for the data providers, the governmental data are provided by various governmental institutions and organizations, and the datasets collected by volunteers are associated with the citizens' initiative *eBird* and the academic project *Stanovanje plus*. On the other hand, of the 21 use cases, 9 are available under the license type CC (CC-BY and CC-BY-SA), 3 are available under an open government license, while the remaining licenses are either private or unspecified. In terms of map type, only 3 maps are dynamic, of which two are interactive and one is viewable. Of the remaining 20 static maps, 18 are interactive. It should also be noted that only one map from the observed use cases is sequential; it shows data about health in Germany in the form of a narrative map. The maps were also analyzed by navigation type. Since a map can contain multiple navigation types, the total number of navigation types exceeds the number of map examples. In total, there are 18 maps with the possibility of spatial navigation, 11 with thematic navigation and 10 with temporal navigation.

From the results obtained, it appears that most of the maps are static and that most of them are maps where the user can customize the display to his own needs. It can also be seen that about half of the maps have thematic and temporal navigation, indicating that there is an awareness of giving the user the possibility to choose the content to be displayed as well as its temporal comparison. In addition, in most cases the maps are based on open government data, which may indicate that this data has better thematic coverage or is more accessible to the general public. It is also interesting to note that only half of the examples (12) are available under "open" licenses, while for the others the license does not clearly indicate how they can be used. It is also clear from the results of the analysis that cartography plays an important role in conveying information from open data, which can be concluded from several use cases. The use case "Identify your watershed and sewer system area", which allows the user to determine the basin, underwater area, and area of the sewer system for address, shows how a map with only three pieces of data can convey information much more clearly than any other means of communication. In addition, the use case "An Interactive Visualisation of NYC Street Trees" which shows the variety and quantity of street trees in the five boroughs, demonstrates that the representation of spatial relationships does not

necessarily have to include traditional cartographic representations, indicating the possibility of adapting cartography to new trends and needs. This is also supported by the fact that five of the use case maps encourage action through visualisation (visual thinking), as opposed to traditional information transfer (visual communication). This confirms the importance of cartography as a mediator between data and the needs of modern society.

Acknowledgments: This research is part of the Twinning Open Data Operational project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 857592.

References

Ansari, B., Barati, M., Martin, E. G. (2022). Enhancing the usability and usefulness of open government data: A comprehensive review of the state of open government data visualization research. Government Information Quarterly 39(1): 101657. https://doi.org/10.1016/j.giq.2021.101657

Chess, C., Johnson, B. B. (2007). Information is not enough. In Moser, S.C.,Dilling, L. (Eds.), Creating a climate for change 223–233. Cambridge University Press. https://doi.org/10.1017/cbo9780

Fish, C. S. (2020). Cartographic content analysis of compelling climate change communication, Cartography and Geographic Information Science 47(6): 492-507. https://doi.org/10.1080/15230406.2020.1774421

Fish, C. S. (2021). Elements of Vivid Cartography, The Cartographic Journal 58(2): 150-166. https://doi.org/10.1080/00087041.2020.1800160

Fish, C. S., Calvert, K. (2017). An Analysis of Interactive Solar Energy Web Maps for Urban Energy Sustainability. Cartographic Perspectives 85. https://doi.org/10.14714/CP85.1372

Kraak, M.-J. (1998). The Cartographic Visualization Process: From Presentation to Exploration. The Cartographic Journal 35(1): 11–15. https://doi.org/10.1179/000870498787074100

Muehlenhaus, I. (2011). Another Goode Method: How to Use Quantitative Content Analysis to Study Variation in Thematic Map Design. Cartographic Perspectives 69: 7–30. https://doi.org/10.14714/CP69.28

Nash, K., Trott, V., Allen, W. (2022). The politics of data visualization and policy making. Convergence 28(1): 3–12. https://doi.org/10.1177/13548565221079156

Ricker, B., Kraak, M-J., Engelhardt, Y. (2020). The power of visualization choices: Different images of patterns in space. In Engebretsen, M., H. Kennedy (Eds.), Data Visualization in Society. Amsterdam: Amsterdam University Press, 2020. https://doi.org/doi 10.5117/9789463722902_ch24

Roth, R.E., Quinn, C., Hart, D. (2015). The Competitive Analysis Method for Evaluating Water Level Visualization Tools. In Brus, J., Vondrakova, A., Vozenilek, V. (Eds.), Modern Trends in Cartography. Lecture Notes in Geoinformation and Cartography. Springer: Cham, Switzerland 241–256.

TOWARDS DEVELOPMENT OF A TOOL FOR THE AUTOMATED ASSESSMENT OF THE SPATIAL ACCURACY OF NATURE OBSERVATION DATASETS

Filip Varga[*1,2], Ana Kuveždić Divjak[3] and Dragica Šalamon[1]

[1] Faculty of Agriculture, University of Zagreb, Croatia
[2] Centre of Excellence for Biodiversity and Molecular Plant Breeding (CroP-BioDiv), Croatia
[3] Faculty of Geodesy, University of Zagreb, Croatia

*correspondence E-mail: fvarga@agr.hr

Keywords: open data; spatial accuracy; big data; toponyms; pipeline concept

1. Introduction

Croatia is one of the most species-rich countries in Europe, thanks largely to its geographical position, where several biogeographical regions overlap, each with its own specific climate, geomorphology, and ecology (Radović et al., 2006). With more than 37,000 taxa within the major taxonomic units, the research on distribution, ecology and conservation, population genetics etc. generates a large amount of spatial data distributed across a large number of taxon-specific databases which are constantly increasing.

Majority of nature obervation databases are a by-product of research projects funded by the government or are set up by non-governmental organisations with little funding. Their data are mostly publicly available, with some exceptions, such as data on endemic and endangered taxa. They often use the help of citizen science initiatives to expand their knowledge of target taxa (Virić Gašparić et al., 2022). Against this background, most nature observation databases, especially the smaller ones, do not have quality control for new records, or only partially apply quality control methods, depending on budget constraints. Input errors related to logical consistency, completeness or duplicate entries are usually mitigated in the initial steps of database architecture development in the form of mandatory fields, input formats and cross-checks with existing records (Dalcin et al., 2012).

Inexperience of researchers in using GPS equipment, such as lack of knowledge of the coordinate reference system (CRS) in which the equipment collects data, writing down coordinates manually instead of using export options on GPS devices often lead to errors in entering records into the database. Errors in the spatial accuracy of records (largely geographical coordinates) are much more difficult to determine. It requires advanced knowledge of GIS (geographical information system) and still takes a lot of time to manually determine whether a particular questionable record has been correctly entered into the database. High-quality spatial data for plant and animal taxa may be the most valuable information for researchers building spatial distribution models to predict the potential

distribution of endangered and endemic taxa, as well as potential risks that invasive species pose to biodiversity, and for government stakeholders involved in decision-making processes for the conservation and sustainable use of species and habitats (Glasnovic et al., 2018; Valencia-Rodríguez et al., 2021).

The aim of this paper was to explore and propose the possibility of developing a pipeline concept for an automated tool to assess the spatial accuracy of nature observation records.

2. Research approach

The first step in our research was a small-scale analysis of the quality of spatial records for Dalmatian pyrethrum (*Tanacetum cinerariifolium* (Trevis.) Sch. Bip.) from the largest Croatian plant database, the Flora Croatica database (FCD) (Nikolić, 2022). The species is endemic and protected by law but has also been cultivated in Dalmatia for a long time as it produces a natural insecticide, pyrethrin, which makes it interesting for both conservation biologists and the agricultural sector. Data on Dalmatian pyrethrum included observational data, herbarium vouchers, published research data and photographs from the natural habitat and was scrapped using Beautiful Soup version 4.9.3. (Python library). The positional accuracy for 906 records in total was checked manually. Points located in water bodies (Dalmatian pyrethrum is a terrestrial species) or outside national boundaries were searched in QGIS version 3.16.9. (QGIS Development Team, 2021).

The second part of the research was to design a pipeline concept suitable for nature observation databases that could be implemented in various open-source software to make it accessible to a wider scientific community. This involved developing theoretical steps for the pipeline detailing the assessment process, exploring different solutions to the specificities of nature observation data, such as the need for geoparsing, i.e., assigning geographic coordinates to free-text descriptions of locations (a commonly used format for describing sampling locations), and finding the best open-source databases with reference toponyms to use in the assessment.

3. Results and discussion

3.1. Preliminary positional accuracy study

The preliminary results showed a relatively high quality of data, with only 4.1% of records classified as spatially inaccurate (1 record was outside national boundaries and 36 records were in water bodies). We had expected the highest positional inaccuracy to occur in nature observation data, which make up the largest share of records (67.8%). For published research data and herbarium specimens, we expected the quality of the specimens to be high, as herbarium curators and researchers check their data extensively before including them in herbarium collections or publishing articles. However, the lowest positional accuracy was found in published research data (10.1% of records are spatially inaccurate), which is alarming.

We are well aware of the limitations of this analysis. Its small scale does not necessarily reflect the quality of the entire database or the state of spatial data in other nature observation databases. The fact that the vast majority of data on Dalmatian pyrethrum in the last two decades (74.5%) was collected using GPS (according to the metadata) may further obscure the

actual quality of the spatial records. If we extrapolate the percentage of low-quality data from this study to the entire FCD, we arrive at a large number of questionable records in terms of spatial quality, namely more than 4,000. Such a large number of records that need to be manually checked would further burden the already understaffed and underfunded institutions that maintain the databases in question.

3.2. Pipeline concept

The first and crucial step in designing this pipeline was to determine the best possible open-source spatial database of Croatian toponyms. The majority of nature observation research data is collected in Croatia, and the location descriptions, including toponyms, are often so small that they cannot be found using global spatial search engines such as Google Maps or OpenStreetMap. For this reason, we strongly recommend using national toponym registers whenever possible. Not only do they contain a large number of records, but they are also standardised and regularly updated. The Register of Geographical Names in the Republic of Croatia consists of 124,018 spatial records and can be downloaded through the WFS service in shapefile format provided by the State Geodetic Administration (State Geodetic Administration, 2021).

The second step we explored was geoparsing. Since nature observation records contain a description of the sampling location in addition to geographical coordinates, and often in free form (e.g., "Dalmatia, Šibenik archipelago, Žirje island, 100 metres from Straža"), we needed to find an efficient way to convert the text identifying a location into a unique geographical reference (from the toponym register), which can then be cross-checked with the sampling coordinates and determine their accuracy. Some of the geoparsing tools we find promising are Mordecai, a full text geoparsing system (Halterman, 2017) and the opencage package for R software (Possenriede et al., 2021). Both are open source and implemented in Python and R respectively, which are known and used worldwide. They can be connected to QGIS and automate the spatial analyses required in this pipeline.

The tool theoretically requires three main components to function properly. The first is the toponym reference dataset in spatial format (shapefile, GeoJSON). Ideally, users should be able to replace the toponym reference dataset with another one so that the tool can be used beyond the Croatian borders. The second component is the sample dataset, which the user provides in tabular form, and which is compared with the toponym reference dataset and checked for spatial accuracy. The third component consists of the software environment in which the whole process takes place. Installation of the tool and use by the user should be easy, and all necessary dependencies should be installed with the tool (Brack et al., 2022). The assessment process described in **Figure 1** is as follows:

1. The user enters the tabular sample dataset into the software environment and defines the CRS of the sample dataset.
2. The tabular sample dataset is converted to spatial format. If necessary, CRS is converted to match the CRS of the toponymic reference dataset.

3.Each record of the sample dataset (description of sampled locations) is geoparsed using the toponym reference dataset. The geographical distance between the sampled location (from the sample dataset) and the reference toponym is calculated.
3.1.If multiple reference toponyms are found in a single record, their names, geographical coordinates, and distance between each reference toponym and sampled location are stored in the sample dataset.
3.2.If no toponyms are found in a single record, this information is stored in the sample dataset and the user receives a notification. The process continues with evaluation of the next record.
4.The process ends when the last record in the sample dataset has been processed.
5.When the assessment of the entire dataset is completed, the user has the option to export:
5.1.the initial sample dataset in tabular form with the addition of the name, geographical coordinates, and distance of the nearest reference toponym to the sampled location for each record.
5.2.The extended sample dataset in tabular form with addition of the name, geographical coordinates and distance of all reference toponyms found for each record. In this dataset, each record is represented with multiple rows depending on the number of reference toponyms found in the location description.
5.3.Initial sample dataset in spatial format (shapefile, GeoJSON)
5.4.Extended sample dataset in spatial format (shapefile, GeoJSON)

Automated assessment of spatial data accuracy could benefit producers (as well as curators) of nature observation data by giving them insight into, and identifying, the number of records that should be assessed more closely for spatial accuracy. The tool could also be of great importance to researchers working on spatial distribution and niche modelling, as it will allow them to identify spatial records for target taxa that they should omit from their datasets to improve the accuracy of the models they develop. The main challenge for the future is the implementation of geoparsing methods for the Croatian language and the evaluation of the success of geoparsing (Gritta et al., 2020).

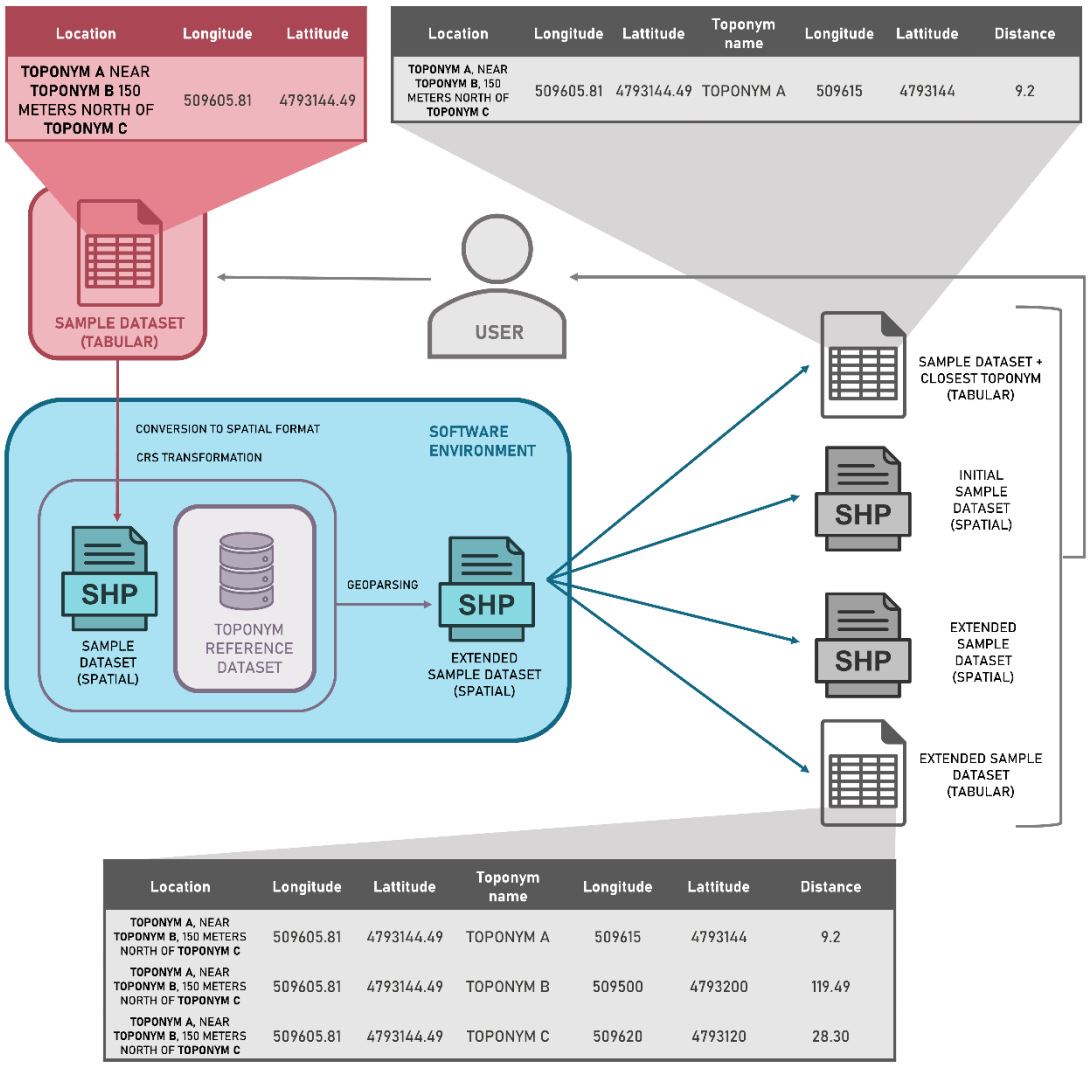

Figure 1. Theoretical pipeline for automated assessment of spatial accuracy of records in nature observation databases.

4. Conclusions and future work

The proposed concept of a tool for the automated assessment of the spatial accuracy of nature observation data sets serves as a design draft that will surely be expanded and modified throughout the implementation process and testing stages. Inclusion of geoinformatics experts and software developers in the next phase of development will provide us with feedback on design flaws and steps necessary to successfully implement this pipeline in a robust assessment tool. After development, variety of nature observation datasets and potentially datasets from other sectors will be tested in order to determine how widely can the tool be used for improvement od spatial records quality on a massive scale.

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

Brack, P., Crowther, P., Soiland-Reyes, S., Owen, S., Lowe, D., R. Williams, A., Groom, Q., Dillen, M., Coppens, F., Gruning, B., Eguinoa, I., Ewels, P., Goble, C. (2022). Ten simple rules for making a software tool workflow-ready. PLOS Computational Biology 18(3): e1009823. https://doi.org/10.1371/journal.pcbi.1009823

Dalcin, E., Silva, L., Cabanillas, C., Loures, M., Monteiro, V., Send Email To Zimbrao, I., Souza, J. (2012). Data Quality Assessment at the Rio de Janeiro Botanical Garden Herbarium Database and Considerations for Data Quality Improvement. In Proceedings of 8th International Conference on Ecological Informatics (ISEI) 1–11. Brasilia, Brasil.

Glasnovic, P., Temunović, M., Lakušić, D., Rakić, T., Grubar, V. B., Surina, B. (2018). Understanding biogeographical patterns in the western Balkan Peninsula using environmental niche modelling and geostatistics in polymorphic *Edraianthus tenuifolius*. AoB PLANTS 10(6). https://doi.org/10.1093/aobpla/ply064

Gritta, M., Pilehvar, M. T., Collier, N. (2020). A pragmatic guide to geoparsing evaluation: Toponyms, Named Entity Recognition and pragmatics. Language Resources and Evaluation 54(3): 683. https://doi.org/10.1007/s10579-019-09475-3

Halterman, A. (2017). Mordecai: Full Text Geoparsing and Event Geocoding. Journal of Open Source Software 2(9): 91. https://doi.org/10.21105/JOSS.00091

Nikolić, T. (Ed.). (2022). Flora Croatica Database. Available at http://hirc.botanic.hr/fcd

Possenriede, D., Sadler, J., Salmon, M. (2021). opencage: Geocode with the OpenCage API. R package version 0.2.2. Available at https://cran.r-project.org/package=opencage

QGIS Development Team. (2021). QGIS Geographic Information System 3.16.9. Open Source Geospatial Foundation Project. Available at http://qgis.osgeo.org

Radović, J., Čivić, K., Topić, R. (Eds.). (2006). Biodiversity of Croatia. Zagreb, Croatia: State Institute for Nature Protection, Ministry of Culture - Republic of Croatia.

State Geodetic Administration. (2021). Register of Geographical Names in the republic of Croatia.

Valencia-Rodríguez, D., Jiménez-Segura, L., Rogéliz, C. A., Parra, J. L. (2021). Ecological niche modeling as an effective tool to predict the distribution of freshwater organisms: The case of the *Sabaleta Brycon henni* (Eigenmann, 1913). PLOS ONE 16(3): e0247876. https://doi.org/10.1371/journal.pone.0247876

Virić Gašparić, H., M. Mikac, K. M., Pajač Živković, I., Krehula, B., Orešković, M., Galešić, M. A., Ninčević, P., Varga, F., Lemić, D. (2022). Firefly Occurrences in Croatia – One Step Closer from Citizen Science to Open Data. Interdisciplinary Description of Complex Systems 20(2): 112–124. https://doi.org/10.7906/INDECS.20.2.4

SESSION VII

OPEN DATA, NEW MEDIA AND CULTURAL HERITAGE

MAPPING THE SCIENTIFIC RESEARCH ON DATA JOURNALISM AND OPEN DATA

Georgios Papageorgiou[1*], **Charalampos Alexopoulos[1]**, Euripidis Loukis[1]

[1]Department of Information and Communications Systems Engineering, University of the Aegean, Greece

*correspondence E-mail: gpapag@aegean.gr

Keywords: Open Data; Data Journalism; Systematic Literature Review

1. Introduction

During the past decade, the available information has increased exponentially from governmental and private sources. On one side governments all over the world are releasing data with the goal to increase their transparency, provide better services to their citizens, and enhance overall efficiency. On the other side, the rise of social media introduced an era of abundant information and instant news propagation like nothing experienced before. Journalists are called to evaluate and process an immense volume of information in this sifting field of direct news and novel emerging data sources(Gray et al., 2012).

In the past journalists were unaccustomed to the use of freely available data in vast quantities and although the use of data, especially in the field of investigative journalism, is not novel the plethora of technological advantages(Calvo-Rubio and Ufarte-Ruiz, 2021; Hassan and Albayari, 2022) provide them with the ability to present a compelling story in a different scale in the data that had been processed and in the potential to visualize and therefore communicate it more compellingly. But these capabilities require specialized technical skills(Veglis and Bratsas, 2017) by the journalists and this can be an important impediment to their utilization of open data.

Data journalism(Bhaskaran et al., 2022; Calvo-Rubio and Ufarte-Ruiz, 2021) and open data(Hurbean et al., 2021; Ismail and Yusof, 2022; Zhang et al., 2018) themselves are well-researched fields in academia but the combination of these two fields is still limited. In this paper, we present a mapping of academic endeavours in the field of data journalism with the use of open data aiming not only to categorize the developments in the field but also to identify benefits and barriers.

2. Methodology

The method used for the analysis of the literature and the identification of topics is a systematic literature review(Okoli, 2015; Xiao and Watson, 2019) (SLR). In the following parts of the chapter, we will analyze the steps of the protocol.

Since this research aims to analyse the progression of publication on data journalism with the use of open data one research question was formulated and refined in greater detail with an analysis framework matrix (**Table 1**).

RQ1: *"How the research literature on data journalism with the use of open data is developing, and what are the most significant benefits and barriers encountered in published papers?"*

Table 1. Analysis framework matrix

Category	Sub-Category	Description
1. Developing of open data research	1.1 Year of the publication	The year when the paper was published
	1.2 Country of the institution	The country where the institution that conducted the research is based
	1.3 Country of research	The country where the research was conducted
2. The topic of the research	2.1 Discipline of research	Use of the Science-Metrix disciplinary classifications
	2.2 Topic of the research	Defines the area explored in the publication
	2.3 Topic of data used	Defines the type of data used in the research (Education, Health, Transportation, Geospatial, etc)
3. Methodologies of the research	3.1 Methodologies for data collection	Survey, Interview, Observation, Literature review, Case study
	3.2 Methodology to analyse data	Statistical analysis, Meta-analysis, Thematic analysis, Content analysis
4. Results of the papers	4.1 Benefits	
	4.2 Barriers	

During the initial search of the research, we defined the keyword query that was used and concluded as "Data journalism" AND "Open Data". The query was used to filter out publications in the title, abstract and keywords on four scientific databases, the databases used are IEEE Xplore, Web of Science, Science Direct, and Scopus.

Initially from these databases were returned 68 publications, the initial exclusion criteria on that step were only the language of publication, and therefore from the 68 publications, we narrowed down the scope of the research to 55, and after the removal of the duplicate documents we ended up in 36.

During the collection of the papers, we were unable to find three, and three more were excluded from the research during the title and abstract filtering therefore the final amount of publications that are examined in this research is 30.

3. Results

The research is still undergoing and therefore in this chapter we will present some preliminary results for the first two categories of the analysis framework. During the research on open data and data journalism, the first step according to the analysis framework is to identify how this particular field is developing. Therefore we collected the data for the year of the publication, the country of the institution that conducted the research country where the research was conducted in the cases where that was applicable.

3.1. Developing open data research

In this part of the research we are examining the development of the field over the previous years and the topography of the publications. In particular, the countries where the institutions conducted these researches are based. In the feature, we will continue with the identification of the counties where these researches were conducted.

3.1.1. Year of the publication

The reason for incorporating the year of publication is to identify the quantitative progression of publications in the field. The interesting observation from the data that are presented in **Figure 1** is that we can see an overall decline in publications since their pick in 2016; Although the open governmental data are increasing(Çaldağ and Gökalp, 2022) the latest bibliometric review for academic publication on open data(Zhang et al., 2018) incorporated data from 2016 and before so we can not be sure that this decline is an emerging trend in the field of open data in general or specific in the use of open data in data journalism.

3.1.2. Country of the institution

The country of the institution is an important factor to identify the countries that are interested in the use of open data in data journalism. As is displayed in **Figure 2** we can observe the dominance of the European Union with 25 publications and the US with 6 publications in the research field. The other countries that do not belong to these groups are the United Kingdom(4), Russia(2), Brazil(1), India(1), Canada(1), Bahrain(1), Switzerland(1), and Palestine(1).

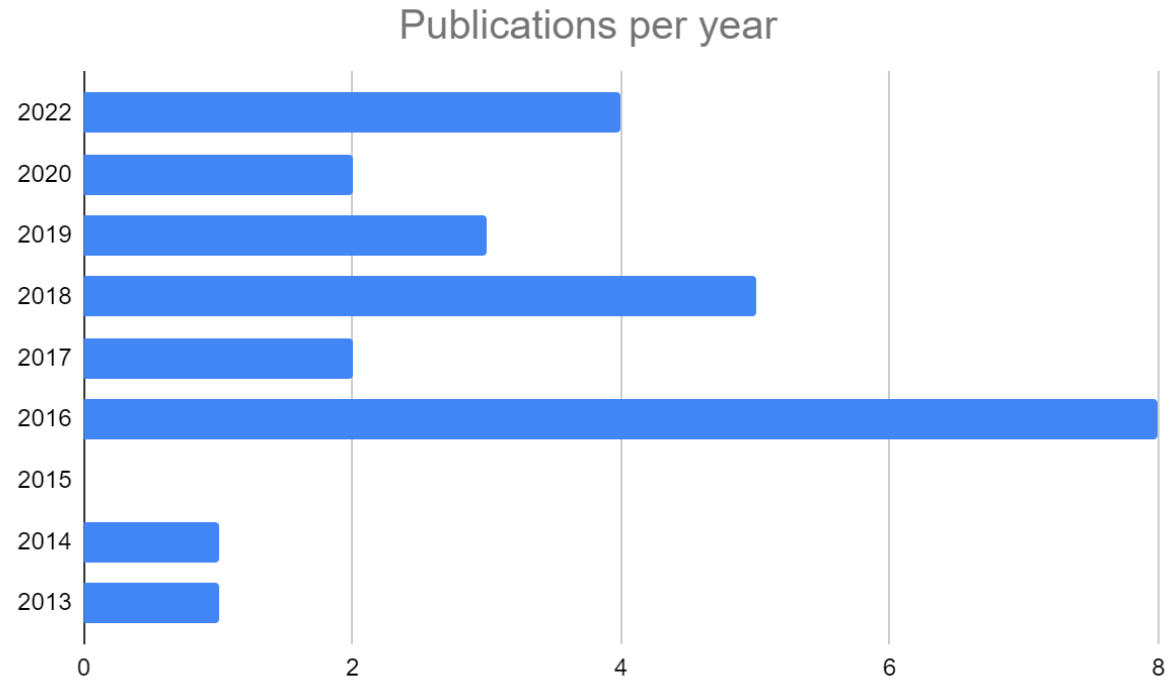

Figure 1. Publications per year

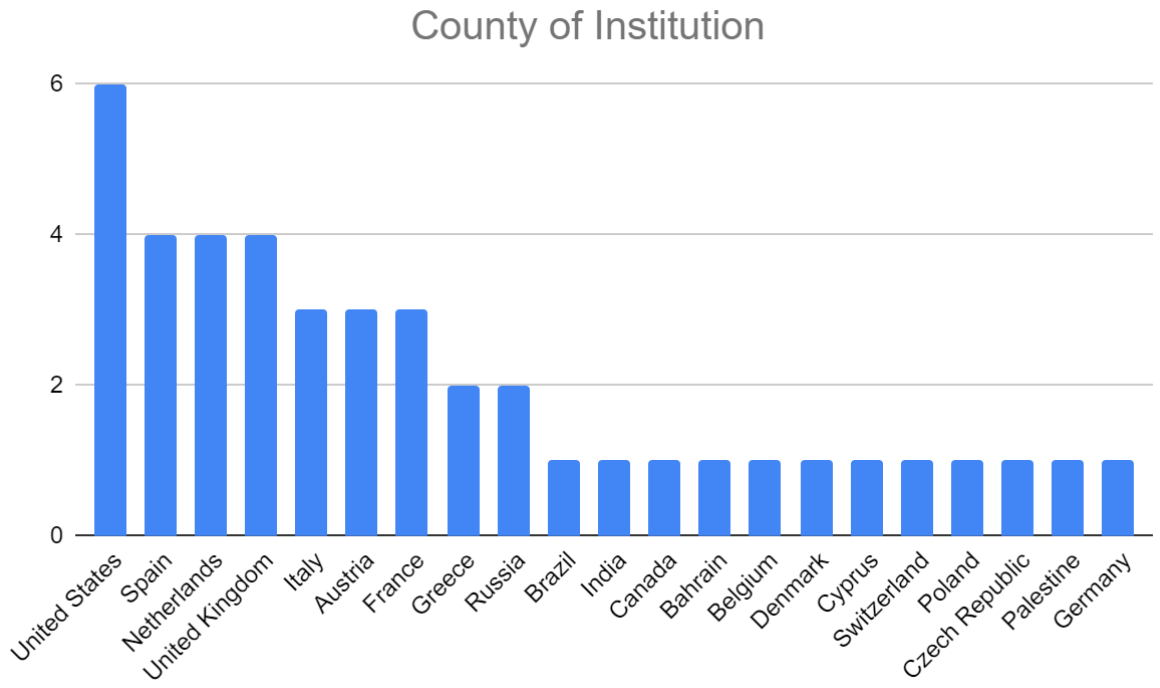

Figure 2. Country of Institution

3.2. Research Topic

In the second part of the research to identify and examine the topic of the research that has been conducted in the field we extracted from the publications the discipline of the research. And we will continue with the purpose that the research, and the topic of the data that has been used where that can be identified.

3.2.1. Research Discipline

The publications were identified to be from different and several academic fields and we had to identify a methodology that can incorporate a wide rand of disciplines. Therefore to identify the disciplinary classifications(Sīle et al., 2021) of this research we used the Science-Metrix classification for journals since it covers the above-mentioned needs of this research. Although we were unable to identify the discipline of three from the selected publications. The Science-Metrix split calcification into three levels and contains 43 categories in the third, we present the finding in that level.

The majority of the publications are in Communication and Media Studies (14) and after that, Information Systems(5), and Artificial Intelligence and Image Processing (4). Overall we can see that the research is balanced between hard and soft science with the technological fields covering 43% of the publications. The data from the classification of the selected publications are presented in **Figure 3** and **Figure 4**.

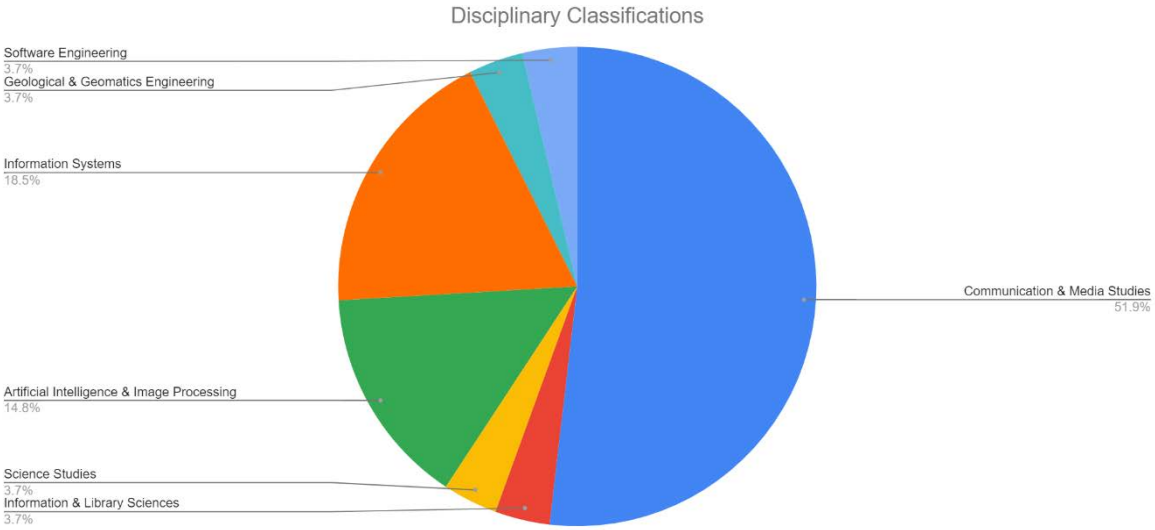

Figure 3. Disciplinary Classification of publications

3.2.2. Research Topic

After we surveyed all the selected publications we compiled a list of topics that were covered by each research, these topics are presented in **Table 2** with here description. Most publications are focused on the technical aspect and especially on the solutions and tools available in the utilisation of open data by data journalists, this trend indicates a gap in the existing technology and the creation of value from the journalistic community. The percentage of topics is consistent with the classification from the journals since the same percentage can be identified in the technological field. One more interesting observation is the broad field that Communication and Media studies contain, in the topic classification this discipline is split into eight categories.

Table 2. Topic Classification and Description

Topic	Description
Legislation comparison	Examine the impact that different legislations have on data available to journalists.
Technological impact	Explore the impact that new technologies have on the field of data journalism.
Technological solution	Present an abstract technical solution that can be implemented also in different cases
Technological tool	Present the use of a particular technical tool.
Cooperation	Investigate the cooperation between data journalists and other professions closely related to the open data field.
Usage of open data	Evaluates the use of open data by data journalists.
Data journalism Practices	Survey the work of data journalists.
User participation	Investigate the participation of the audience in the creation of open data.
Communication models	Examine different communication models for the dissemination of news.
Education	Identify the educational needs and opportunities in data journalism.
Digital skills	Examine the digital skills needed for the utilisation of open data.

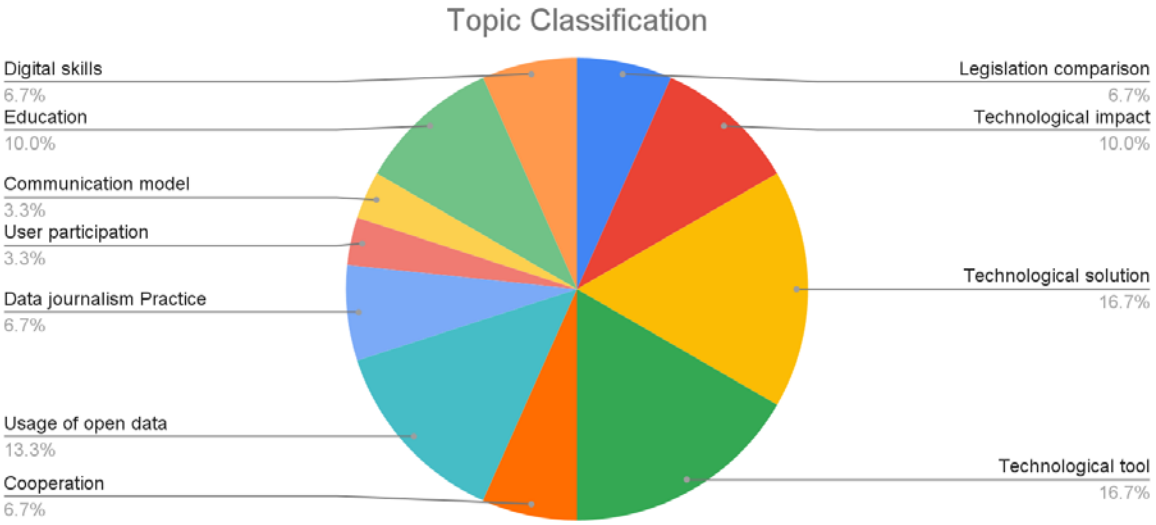

Figure 4. Topic Classification per publication

4. Conclusion

In the exploration of the use of open data in data journalism we are conducting a systematic literature review of the published research on the topic. Our results until now are showing a decrease in the yearly publications and the dominance of the European Union and the US in the field with the majority of the research publication being from institutions in these two areas. Also in the topics that have been researched, we discovered that are balanced between social and technical, publications.

Since the research is undergoing we still have complete the analysis of the publication to determine the countries where these researches were conducted. Furthermore, the methodologies that have been used for the collection and analysis of the data have to be ascertained, and finally, the benefits and barriers of these publications have to be discovered.

Acknowledgements: We acknowledge the financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 955569. The research was partially funded by TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

References

Bhaskaran, H., Kashyap, G., Mishra, H. (2022). Teaching Data Journalism: A Systematic Review. Journalism Practice: 1–22. https://doi.org/10.1080/17512786.2022.2044888

Çaldağ, M. T., Gökalp, E. (2022). The maturity of open government data maturity: A multivocal literature review. Aslib Journal of Information Management 74(6): 1007-1030. https://doi.org/10.1108/AJIM-11-2021-0354

Calvo-Rubio, L.-M., Ufarte-Ruiz, M.-J. (2021). Artificial intelligence and journalism: Systematic review of scientific production in Web of Science and Scopus (2008-2019). Communication & Society 34(2): 159–176. https://doi.org/10.15581/003.34.2.159-176

Gray, J., Bounegru, L., Chambers, L., European Journalism Centre, Open Knowledge Foundation (Eds.). (2012). The data journalism handbook (1st ed). O'Reilly Media.

Hassan, A., Albayari, A. (2022). The Usage of Artificial Intelligence in Journalism. In Hamdan, A., Harraf, A., Arora, P., Alareeni, B., Khamis Hamdan, R. (Eds.) Future of Organizations and Work After the 4th Industrial

Revolution. Studies in Computational Intelligence, vol 1037. Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-99000-8_10

Hurbean, L., Danaiata, D., Militaru, F., Dodea, A.-M., Negovan, A.-M. (2021). Open Data Based Machine Learning Applications in Smart Cities: A Systematic Literature Review. Electronics 10(23): 2997. https://doi.org/10.3390/electronics10232997

Ismail, N., Yusof, U. K. (2022). Recent trends of machine learning predictions using open data: a systematic review. Journal of Information and Communication Technology 21(3). https://doi.org/10.32890/jict2022.21.3.3

Okoli, C. (2015). A Guide to Conducting a Standalone Systematic Literature Review. Communications of the Association for Information Systems 37. https://doi.org/10.17705/1CAIS.03743

Sīle, L., Guns, R., Vandermoere, F., Sivertsen, G., Engels, T. C. E. (2021). Tracing the context in disciplinary classifications: A bibliometric pairwise comparison of five classifications of journals in the social sciences and humanities. Quantitative Science Studies 2(1): 65–88. https://doi.org/10.1162/qss_a_00110

Veglis, A., Bratsas, C. (2017). Reporters in the age of data journalism. Journal of applied journalism & media studies 6(2): 225–244. https://doi.org/10.1386/ajms.6.2.225_1

Xiao, Y., Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. Journal of Planning Education and Research 39(1): 93–112. https://doi.org/10.1177/0739456X17723971

Zhang, Y., Hua, W., Yuan, S. (2018). Mapping the scientific research on open data: A bibliometric review. Learned Publishing 31(2): 95–106. https://doi.org/10.1002/leap.1110

TWITTER SENTIMENT ANALYSIS ON THE IMPLEMENTATION OF ONE DATA INDONESIA WITH SEMI SUPERVISED SVM

Nissa Silvianna Devi Nur Afni[1*], Yuliagnis Transver Wijaya[2]

[1] Badan Pusat Statistik, Indonesia
[2] Politeknik Statistika STIS, Indonesia

*correspondence E-mail: nissasilviannadna@gmail.com

Keywords: One Data Indonesia; Semi-Supervised; SVM; Resampling; TF-IDF

Data is a new source of wealth. Data sovereignty must be realized immediately (Presiden RI, 2019). The quality data is a reference for policy formulation, especially in planning, monitoring and evaluating programs. The use of quality data in policy formulation is very useful to achieve the targets properly so the development goals can be achieved effectively (Badan Pusat Statistik, 2010). Data quality plays an important role in the preparation of national-level work plans to achieve the Sustainable Development Goals (SDGs) (Sekretariat Nasional Open Government Indonesia, 2020).

Quality is a form of measurement of the extent to which a set of characteristics inherent in an object meets the requirements. In statistical organization, the objects are statistical output, statistical process, institutional environment, and statistical system. The quality of statistical output is very influential in the quality management framework because all components lead to statistical output (United Nations, 2019). To obtain quality and easily accessible data, the Indonesian government is improving data governance through the implementation of One Data Indonesia. One Data Indonesia (ODI) is a government data management policy to support development planning, implementation, evaluation, and control (Peraturan Presiden No.39 Tahun 2019 Tentang Satu Data Indonesia, 2019).

The implementation of ODI needs to notice the principle of One Data Indonesia. Data generated by data producers must meet data standards, have metadata, meet data interoperability rules, and use reference codes and/or master data. With ODI, it is hoped that any data published through the ODI portal can be used freely, in an open format, and can be used and redistributed by anyone without conditions other than citing the source and data owner (Peraturan Presiden No.39 Tahun 2019 Tentang Satu Data Indonesia, 2019). Thus, the use of government data is not only limited to internal use between agencies, but also as a form of fulfilling public data needs for the community (Indrajit, 2018).

The tangible implementation of ODI started in 2017 (Media Indonesia, 2016). This was conveyed by the Minister of National Development Planning that the use of one data was started for program

planning for the 2017 fiscal year. The implementation of one data in Indonesia was started to be applied by ministries/agencies by integrating the data into one comprehensive data.

In its implementation, ODI still faces obstacles and challenges, such as inconsistent and scattered data in various public institutions with difficult access, lack of coordination between data-owning institutions, and non-standardized data. Important components that must be maximized in the implementation of ODI are strengthening data management and governance, publication and dissemination of open data through portals, as well as involvement and sustainability of data users (Islami, 2021). As one of the main components in ODI, the opinion of data users on the implementation of ODI is important to be researched. One of the data sources that can be used to analyze opinions is Twitter data. Twitter provides services to its users to read and write opinions. In addition, Twitter users in Indonesia are also quite widespread. By the 191.4 million Indonesians who actively use social media, 63.6% of them use Twitter (Global Digital Insights, 2021).

Sentiments contained in the opinions of Twitter users can be analyzed into useful information. Sentiment is the center of various communication studies where negativity and polarization have a significant influence on economic conditions, social life, politics, to development issues that are integrated with information technology. The dynamics of interaction between social media users is considered representative so that it becomes important data to be analyzed as development recommendations (Singgalen, 2021).

Sentiment analysis is a computational study of people's opinions, feelings, emotions, and attitudes towards an entity such as products, services, issues, events, topics, and attributes. Sentiment analysis is one of the most active studies currently in the study of natural language processing, data mining, information retrieval, and web mining (Antonio et al., 2017). Sentiment analysis can be a more effective and cost-effective alternative to surveys (Wei et al., 2021).

Sentiment analysis with machine learning is divided into 3 categories, they are unsupervised learning, semi-supervised learning, and supervised learning. The three approaches have their respective advantages. The advantages of unsupervised learning include saving time and money because there is no need to labeled the data. Supervised learning requires more energy to label the data, but the model that is formed usually has better performance than other methods (Ligthart et al., 2021).

The semi-supervised learning expected to reduce the shortcomings of the two previous approaches by collaborating them. Semi-supervised learning uses unlabeled data supplemented by labeled data to form a classification model. Semi-supervised learning is commonly used for sentiment analysis with large amounts of unlabeled data. This model produces quite good results when the dataset is very large (Ligthart et al., 2021). Semi-supervised learning is popularly applied to sentiment analysis of Twitter data with a lot of unlabeled data (Da Silva et al., 2016). Based on the data held by the researcher, the semi-supervised learning is considered to be the most suitable for this research.

Currently, semi-supervised learning has been developed. Several semi-supervised learning based on research by Jamshid Bagherzadeh in 2018 include self-training methods, generative methods, co-training methods, margin-based methods, graph-based methods, and semi-supervised boosting. Pseudo-labeling is a technique in self-training that can be applied to both labeled and unlabeled data (Fudholi and Juwairi, 2009). Yulan He and Deyu Zhou mentioned that the self-training with

pseudo-labeling is a simple, robust, and does not need to focus too much on tuning parameters (He and Zhou, 2011). In addition, Lee Dong-Hyun in his research mentions that the pseudo-labeling method is a simple and efficient method (Dong-Hyun Lee, 2013).

Pseudo-labeling techniques can be used in machine learning methods such as Naïve-Bayes, Neural Network, etc. Ligthart et al. in 2021 compared three machine learning, namely Support Vector Machine (SVM), Naïve Bayes, and Random Forest with a self-training. Naïve Bayes produces the best accuracy with the highest accuracy of 94%. Then Chandani et al. do a comparison between SVM, Naïve Bayes and Artificial Neural Network where SVM produces the best accuracy (Chandani et al., 2015). Furthermore, Ipmawati et al. perform comparisons between SVM, Naïve Bayes and K-Nearest Neighbor (K-NN). In this study SVM also produced the best accuracy compared to the other two algorithms (Ipmawati et al., 2017).

This research was begun with data collection. Data were obtained by scraping data from January 1, 2017 to December 31, 2021 using the library "Twint" based on 8 keywords. The keyword is determined by researchers based on words that have a relationship with One Data Indonesia. The data collection results 21,286 unique data with detailed data can be seen in **Table 1**.

Table 1. The results of scraping Twitter data.

Keyword	Total
@datagoid	1,311
data terbuka	7,622
data.go.id	136
integrasi data	948
manajemen data	3,406
pengelolaan data	3,170
data governance	927
satu data Indonesia	3,855
Total	**21,286**

Data from Twitter is then labeled manually. Tweets are classified into positive (1), negative (-1) and neutral (0). Positive and negative labels are given to their tendency towards SDI implementation. While the neutral label is given to tweets that do not show a positive or negative trend, such as tweets containing information, notifications, and tweets that are not related to the implementation of one data Indonesian, such as tweets about internships, webinars, etc. There is 2,778 data used in manual labeling which can be categorized in 323 in positive, 2,153 in neutral, and 302 in negative. Data from manual labeling categorized in 2,778 data with positive categories as many as 323, neutral 2,153, negative 302. Then the text preprocessing to remove noise in the data.

Classification with a semi-supervised applied by technique pseudo-labeling. The method used is SVM with 5-fold cross validation. In each modeling, weighting is done with TF-IDF and resampling with SMOTE. The modeling in this study was carried out 2 times. Data is divided into several sets as seen in **Table 2**.

Table 2. The distribution of dataset for modeling.

Dataset Name	Ammount of Data
Labeled data	2,729
Unlabeled Data I (25%)	4,076
Unlabeled Data II (75%)	12,227
Total	**19,032**

The first modeling is applied to manually labeled data. The data is divided into 3 data sets, namely training, testing, and validation. The model is formed with training and predictions are made on data testing and validation. Resampling is only done on training data. First model produces an average accuracy of 0.937931 with a standard deviation 0.008062. The model is used to make predictions on 25% of the non-labeled data. Prediction results are presented in **Figure 1**.

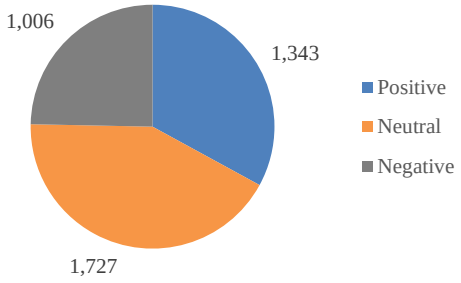

Figure 1. Prediction results 25% of unlabeled data in first modeling.

The classification results are assessed using a confusion matrix to see the model's performance in classifying tweet data. The first model produces an accuracy value obtained by testing the model with data validation of 82.03% data testing of 80.67%. The accuracy value is quite consistent, indicating that the first model is quite good at classifying. Based on the **Figure 2**, it can be concluded that the model predicts well for the neutral category and quite well for the positive and negative categories.

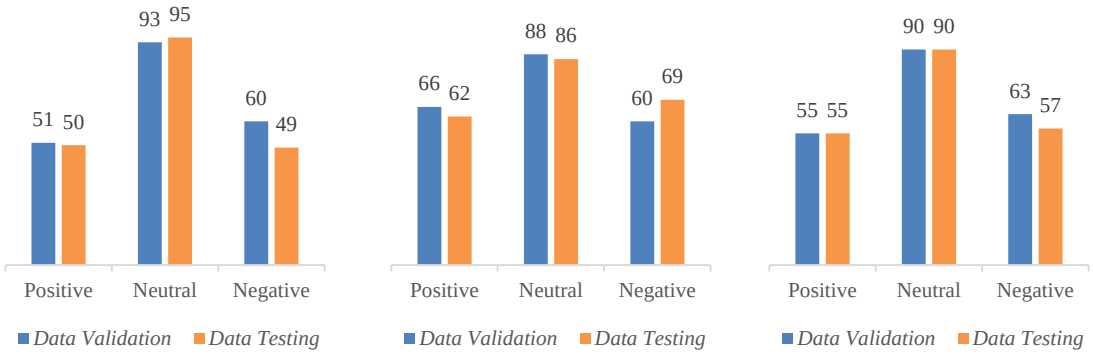

Figure 2. Modeling Precision, Recall, and F1-Score of the first moedling.

The second modeling is applied to the combined data of training data and predicted data from the first model. The modeling is done using the same technique as the first modeling. The second model produces an average accuracy of 0.916902 with a standard deviation of 0.011869. The model is used to make predictions on 75% of the unlabeled data. Prediction results are presented in **Figure 3**.

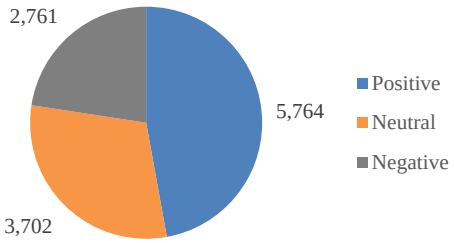

Figure 3. Prediction results 25% of unlabeled data in second modeling.

The second model produces an accuracy value of 79.70% for testing with data validation and 82.25% with data testing. The accuracy value in the second model is also quite consistent which shows that the second model is also quite good at classifying. Based on the **Figure 4**, it can be concluded that the model can predict well in the neutral category and quite well for the positive and neutral categories. Furthermore, the model tends to be better at predicting the positive and negative categories, but slightly worse for the neutral category.

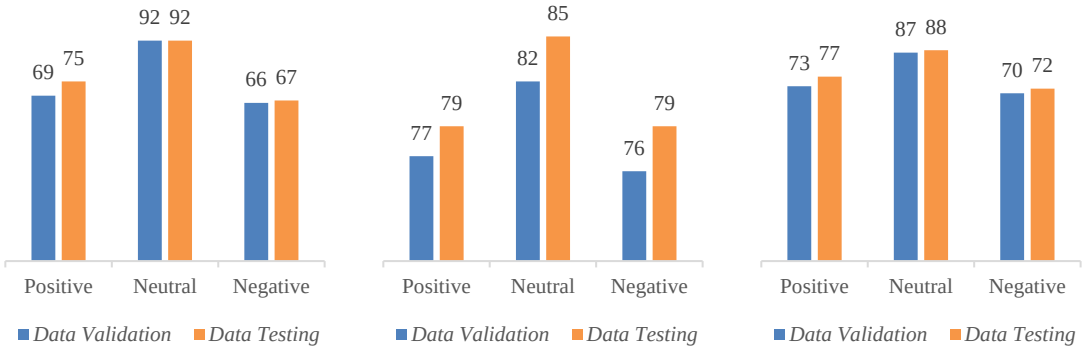

Figure 4. Modeling Precision, Recall, and F1-Score of the second modeling.

In this research, the researchers also make a comparison of the performance of semi-supervised learning techniques pseudo-labeling on Support Vector Machine (SVM), Naïve Bayes (NB) and K-Nearest Neighbors (KNN) algorithms to see how well our methods in clasifying the tweets. The performance to be compared is the result of cross validation and the accuracy value. First comparison was made for model 1 using data labeled as training data. Based on the cross validation shown in **Figure 5**, SVM has the highest accuracy value followed by Naïve Bayes, and K-Nearest Neighbors. Based on the **Figure 6**, it can be seen that SVM has the lowest standard deviation value. This shows that SVM is the best model in terms of cross validation.

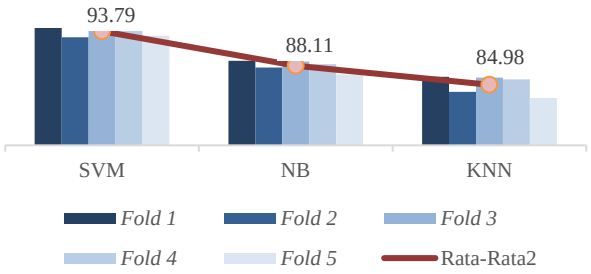

Figure 5. Comparison of cross validation in first modeling.

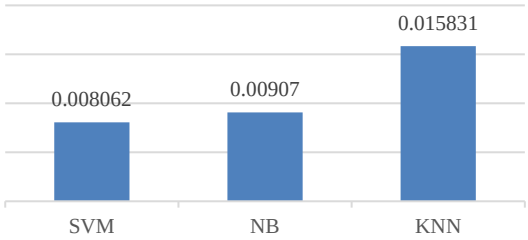

Figure 6. Comparison of the standard deviation in first modeling.

The next comparison is carried out in second modeling, where the training data used is a combination of the training data used in first modeling and pseudo-labed data (predicted data from first modeling). Based on the cross validation shown in **Figure 7**, SVM has the highest accuracy value followed by NB and KNN. Based on **Figure 8**, it can be seen that SVM has the lowest standard deviation value. Furthermore, KNN has a standard deviation of accuracy that is not much different from SVM then followed by NB with the highest standard deviation of accuracy. Based on modeling2, SVM still produces the best model in terms of cross validation.

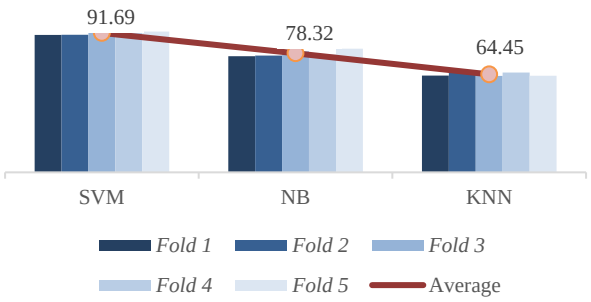

Figure 7. Comparison of cross validation in second modeling.

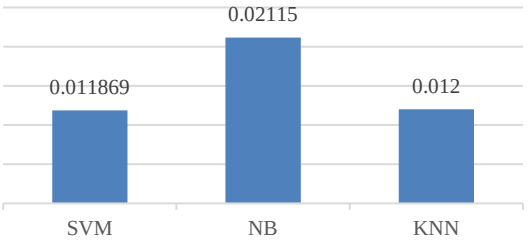

Figure 8. Comparison of the standard deviation in second modeling.

The first comparison was made for first modeling using data labeled as training data. Based on the **Figure 9**, it can be concluded that SVM has the best accuracy in first modeling compared to the other two algorithms. While KNN has the lowest accuracy and NB is slightly better.

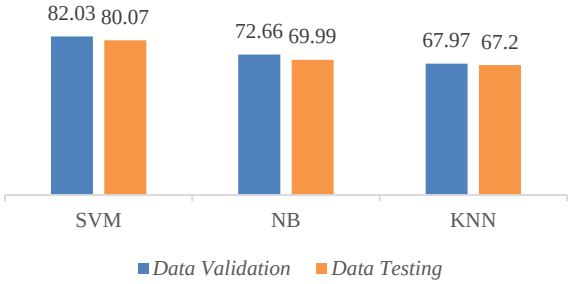

Figure 9. Comparison of accuracy in first modeling.

Furthermore, second modeling is carried out using training data in the form of a combination of training data for first modeling with pseudo-labeled. Prior to modeling, resampling was also carried out in the second modeling. Based on **Figure 10**, it can be concluded that SVM has the best accuracy in first modeling compared to the other two algorithms. While KNN has the lowest accuracy and NB is almost close to SVM.

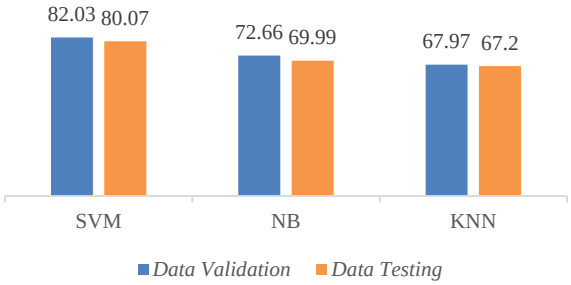

Figure 10. Comparison of accuracy in second modeling.

The results of the sentiment classification show that the sentiments of Twitter users regarding the implementation of One Data Indonesia tend to be positive and neutral. Positive sentiments of Twitter users tend to be related to systems, integration, government and governance. Meanwhile, on the negative sentiment, some Twitter users also discussed systems, governance, and government. This shows the lack of stability in the system, governance, and government so that different sentiments emerge.

The word that tends to be used a lot in positive sentiment is the word "integration". This shows that the opinion of Twitter users regarding integration in Indonesia tends to be positive. While the word that tends to be used a lot in negative sentiments is "valid". This indicates that Twitter users' sentiment towards data validity in Indonesia tends to be bad.

Sentiment classification of tweets using the SVM algorithm results in an accuracy testing data of 80.67%. Then the classification with semi-supervised SVM produces an accuracy of 82.25%. The increase in accuracy indicates that the addition of pseudo-labeled can improve accuracy. In addition, semi-supervised SVM proved to have better performance than semi-supervised Naïve Bayes, and semi-supervised K-Nearest Neighbor.

References

Alexopoulos, C., Loukis, E., Mouzakitis, S., Petychakis, M., Charalabidis, Y. (2018). Analysing the Characteristics of Open Government Data Sources in Greece. Journal of the Knowledge Economy 9(3): 721-753. https://doi.org/10.1007/s13132-015-0298-8

Antonio, E. C., Das, D., Bandyopadhyay, S., Feraco, A. (2017). Sentiment Resources: Lexicons and Datasets. In A Practical Guide to Sentiment Analysis 85-106. https://doi.org/10.1007/978-3-319-55394-8_5

Badan Pusat Statistik. (2010). Rencana Strategis Badan Pusat Statistik.

Barbero, M., Bartz, K., Linz, F., Mauritz, S., Wauters, P., Chrzanowski, P., Osimo, D. (2018). Study to support the review of Directive 2003/98/EC on the re-use of public sector information. Available at https://ec.europa.eu/digital-single-market/en/news/impact-assessment-support-study-re-vision-public-sector-information-directive

Chandani, V., Wahono, R. S., Purwanto. (2015). Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film. Journal of Intelligent Systems 1(1): 56–60.

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., Ferro, E. (2018). Open data interoperability. In The World of Open Data 75–93. https://doi.org/10.1007/978-3-319-90850-2_5

Da Silva, N. F. F., Coletta, L. F. S., Hruschka, E. R., Hruschka, E. R. (2016). Using unsupervised information to improve semi-supervised tweet sentiment classification. Information Sciences 355–356: 348–365. https://doi.org/10.1016/j.ins.2016.02.002

Davies, T., Walker, S. B., Rubinstein, M., Perini, F. (2019). The State of Open Data Histories and Horizons.

Dong-Hyun Lee. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. ICML 2013 Workshop: Challenges in Representation Learning, July 2013, 1–6.

EIF. (2017). European Interoperability Framework. In White Pages. Available at https://doi.org/10.2799/78681

European Commission. (2020a). eGovernment Benchmark 2020: eGovernment that works for the people.

European Commission. (2020b). Public sector information, open data and big data (RP2020).

Fudholi, D. H., Juwairi, K. P. (2009). Pemanfaatan Teknik Semi-Supervised Learning untuk Klasifikasi Dokumen Medis. Universitas Islam Indonesia.

Global Digital Insights. (2021). Digital 2021 Indonesia - We Are Social. In Global Digital Insights.

He, Y., Zhou, D. (2011). Self-training from labeled features for sentiment analysis. Information Processing and Management 47(4): 606–616. https://doi.org/10.1016/j.ipm.2010.11.003

Indrajit, A. (2018). One Data Indonesia to Support the Implementation of Open Data in Indonesia. In Open Data Exposed. T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-261-3_13

Ipmawati, J., Kusrini, Luthfi, E. T. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. Indonesian Journal on Networking and Security 6(1): 28–36.

Islami, M. J. (2021). Implementasi Satu Data Indonesia: Tantangan dan *Critical Success Factors* CSFs). Jurnal Komunika: Jurnal Komunikasi, Media Dan Informatika 10(1): 13. https://doi.org/10.31504/komunika.v10i1.3750

Joinup Europe. (2022). Interoperability layers Translate. Interoperability Solutions. Available at https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/3-interoperability-layers

Ligthart, A., Catal, C., Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. Artificial Intelligence Review 54(7): 4997-5053. https://doi.org/10.1007/s10462-021-09973-3

Media Indonesia. (2016). Penerapan Satu Data dari BPS Mulai 2017. Media Indonesia. Available at https://mediaindonesia.com/ekonomi/42866/penerapan-satu-data-dari-bps-mulai-2017

Mila Gascó-Hernández, Erika G. Martin, Luigi Reggi, Sunyoung Pyo, L. F. L.-R. (2018). Promoting the use of open government data: Cases of training and engagement. Government Information Quarterly 35(2): 233–242. https://doi.org/10.1016/j.giq.2018.01.003

Peraturan Presiden No.39 (2019). Tahun 2019 tentang Satu Data Indonesia, Peraturan Presiden 1.

Presiden RI, J. W. (2019). Pidato Kenegaraan.

Renee E. Sieber, P. A. J. (2015). Civic open data at a crossroads: Dominant models and current challenges. Government Information Quarterly 32(3): 308–315. https://doi.org/10.1016/j.giq.2015.05.003

Sekretariat Nasional Open Government Indonesia. (2020). Rencana Aksi Keterbukaan Pemerintah Indonesia 2018-2020.

Singgalen, Y. A. (2021). Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Sistematic Literature Review. Journal of Information Systems and Informatics 3(2): 278–302. https://doi.org/10.33557/journalisi.v3i2.125

Umbrich, J., Neumaier, S., Polleres, A. (2015). Quality assessment and evolution of Open Data portals. Proceedings - 2015 International Conference on Future Internet of Things and Cloud, FiCloud 2015 and 2015 International Conference on Open and Big Data, OBD 2015, 404–411. https://doi.org/10.1109/FiCloud.2015.82

United Nations. (2019). United Nations National Quality Assurance Frameworks Manual for Official Statistics. In United Nations National Quality Assurance Frameworks Manual for Official Statistics. United Nations Publication. https://doi.org/10.18356/1695ffd8-en

van Loenen, B. (2018). Towards a User-Oriented Open Data Strategy. In Open Data Exposed 33–53. https://doi.org/10.1007/978-94-6265-261-3_3

van Loenen, B., Zuiderwijk, A., Vancauwenberghe, G., Lopez-Pellicer, F. J., Mulder, I., Alexopoulos, C., Magnussen, R., Saddiqa, M., Dulong de Rosnay, M., Crompvoets, J., Polini, A., Re, B., Casiano Flores, C. (2021). Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. JeDEM - EJournal of EDemocracy and Open Government 13(2): 1–27. https://doi.org/10.29379/jedem.v13i2.644

Wei, Y., Gong, P., Zhang, J., Wang, L. (2021). Exploring public opinions on climate change policy in "Big Data Era"—A case study of the European Union Emission Trading System (EU-ETS) based on Twitter. Energy Policy 158: 112559. https://doi.org/10.1016/j.enpol.2021.112559

OPTIMISATION OF MEDIA REPORTING IN THE CONTEXT OF OPEN DATA DEVELOPMENT

Dora Gelo Čolić[1*]

[1] Faculty of Humanities and Social Sciences, University of Zagreb, Croatia

*correspondence E-mail: doragelo@gmail.com

Keywords: optimisation; data verifiability; media reporting; relevance; open data; public interest

1. Introduction

The right measure of media reporting on sensitive issues related to national security would be the optimal amount of information to integrate the essential tension between two contradictory demands. In this case we are talking about the demand for openness, in other words, the demand to give access to information that is critical for citizens, or, probably the best description, information of public interest. On the other side is the demand for secrecy with regard to national security.

The content of the concept of public interest could be raised as a debatable question. However, it is unlikely that subtle differences in the perception of public interest would significantly erase the importance of factors such as truthfulness of information, relevance of information, clarity and coherence of reporting. Quoting McQuail (2015):

> "…there is the idea of 'public interest' as in the chapter title, used widely and diversely, sometimes misused and escaping an agreed definition except in very general terms (as a notion of the welfare of the public or society as a whole). However, Jay Blumler did suggest (1998) a fairly clear view of what it would mean, in terms of three key features: it refers to a vision of what is good for the many as instituted by some form of legitimate democratic authority (as are certain other functions in society, such as government or the justice system) and having an element of public responsibility as a consequence. It is an idea that requires a long time span, which takes into account the long-term needs and effects, thus of future as well as present generations. In addition, the notion of a public interest has to recognize conflicts of interest and of perspective in implementation; it must permit compromise and adaptation to the realities of the time and place."

McQuail sees media theory as a complex structure of socio-political and philosophical principles, which organises ideas on the relationship between the media and society. As one of the few theories, in his classification we find the normative media theory. It deals with the issue

of what the media should do rather than what they are really doing. Despite the fact that McQuail categorises theory in a few basic varieties, he adds that in reality there are no clean models; rather, what exists in a concrete society is a model combining theoretical elements and media types. He also believes that basic principles of media activity can be isolated as independence, diversity or pluralism, information quality and the preservation of social and cultural order. He warns that some of these principles are conflicting, but that one of the aims of regulation is the very management of tensions and settlement of conflicts (McQuail, 2005). The conflict in focus in this research is between secrecy and openness, which we see as something that the national security system is entitled to do vs. the demand of the public as part of the norm of a democratic society. But it is just one of the many characteristics of this specific relationship between the system and the media. As stated by authors Peter Gill and Mark Phythian (2018) in their book "Intelligence in an Insecure World", a part of their relationship can also be marked by a combination of dependence, manipulation, support and praise. These appearances are also directly connected to the possible negation of basic principles like independence and information quality.

As a wide variety of problems in science, engineering etc. can be posed as optimisation problems (Jain et al., 1996), the goal of a formed model, and hopefully an optimisation algorithm in the future, is to calculate that right measure, naturally following the results from the conducted research process, and to optimise information. Secondly, an alternative to optimising every piece of information (in this research information equals sentence) is to optimise the whole media article by picking only sentences that fall in the range of calculated values that are determined as those that legitimise publishing based on the scope and goal of the research.

2. Methodology

The methodology used in this research is mostly based on Florian Kohlbacher design. Firstly, a theoretical analysis was conducted. Afterwards, in the second phase, results of the theoretical analysis were applied to a certain amount of units. When we speak about units we mean sentences from the chosen media article. In this particular research we examined 34 sentences. We use structure analysis, syntax analysis, grammatical analysis and narrative analysis.

An empirical analysis after the theoretical analysis filtered the applied theoretical results. That's how we obtained the matrix as the meeting point of theoretical terms and their empirical usage. By picking patterns, the table was created and it contains formulas expressing connections between both empirical and theoretical examined elements.

In the third phase of research, the analysis was done based on the produced table and the results were recorded. Considering the overall goal of the research, many more units have to be included in order to achieve a bigger number of representative examples.

3. Theme framework

A theme framework needs to be defined in order to create a scope for the researching of fundamental theoretical terms in information sciences. Fundamental theoretical terms, on the one hand, are relevance, informativeness and usefulness, and on the other hand, they are data

and information.The theme framework is related to the national security concept but in its specific non-military aspects, like the inner structural weakness of the system, the vulnerability of the institution, the vulnerability of the value system related to political stability, and social cohesion (Buzan, 1983). Together with the aforementioned points, we are including situations when hybrid threat activity, by applying combinations of tools, targets a state in the legal, diplomatic, information and political domains. Then we talk about each tool targeting one or multiple domains, or the interface between them, by creating or exploiting a vulnerability or taking advantage of an opportunity (Giannopoulos, 2021). We could put all of the definitions under the same umbrella of finding a balance between democratic standards and registering hybrid threats in timely manner.

4. Theoretical analysis

The general definition of information (GDI) is an operational standard that defines the information as an item that 1) is composed of one piece of data or more data that 2) is well composed and 3) has meaning.

By well-composed data, we mean data that are composed in accordance with the syntax of a certain system, whereby syntax is seen as a term in a broader sense than linguistics, which includes the determination of form, construction, composition and structure (Floridi, 2011). The fundamental nature of the data is where we see x different to y, where x and y are two uninterpreted variables, or in other words, the fundamental nature of data is exhibiting the anomaly (Floridi, 2009).

"The term 'information' is often used to represent a variety of things, events or expressions. In other words the term 'information' cannot be interpreted as the things, events or expressions themselves unless it is situated within a context and is understood meaningfully" (Ma, 2012). Considering that, the research is done in a specific context to examine theory. Therefore, its results are to be interpreted only as such.

Over time, relevance was explored in different directions, but it is possible to summarise it into four approaches to the nature of relevance. It was found that the following emerged: systemic, communicative, situational and psychological, and we can add to these a fifth or interactive framework based on a layered interaction model of information searching, where interactions include levels or layers. It is considered that there is not only one relevance but an interdependent system of relevances, which dynamically interact with each other within and between different layers or levels and adapt when necessary. A categorisation of the manifestation of relevance was proposed and it was cross-referenced with the system of relevance. It is considered that relevance has a set of general features that characterise its nature, which includes:

a) relation
b) intention
c) context
d) inference by involving an assessment of a relation, often an advanced assessment of the success or degree of enhancement of a particular relation, such as an assessment of some information sought in relation to a context-directed intention

e) interaction meaning an inference is achieved as a dynamic, interactive process in which interpretations of other features can change, depending on understanding (Saračević, 2006).

Following this, in this research we consider relevance as the element strongly related to the characteristic of data, of which the particular information consists in the specific context. We find it is an unchangeable characteristic in the perspective of time.

$$R = (is + pr) \times ST$$

where

R - relevance
ST - truthlikeness
is – average correctness of data
pr – average preciseness of data

Counting on the fact that average preciseness and average correctnes are not chanegable, the only change we could expect in this is that of truthlikeness. We didn't include in our work the concept of truth because we strongly believe that it would limit the scope of the research and its usefulness. But we use the term truthlikeness. It is a concept that we observe in the following framework.

As there might be a betweenness relationship amongst worlds, or even a fully-fledged distance metric, we can examine what is closer to the truth. The essence of the likeness approach is that the truthlikeness of a proposition is somehow dependent on the likeness between worlds. Graham discusses three main problems for any concrete proposal within the likeness approach:

a) an account of likeness between states of affairs –what does this consist of and how can it be analysed or defined?
b) the dependence of the truthlikeness of a proposition on the likeness of worlds in its range to the actual world: what is the correct function?
c) "translation variance" or "framework dependence" of judgements of likeness and of truthlikeness (Oddie, 2016).

The truthfulness that is a changing measure on this scale is actually the degree of realism of a particular claim or information, its proximity to the real state of affairs. The problem is determining the proximity to truth/reality/information (Floridi, 2011).

Following what was mentioned earlier in the research, we concluded that truthlikeness for the purpose of this research is the proportion of verifiable data from the total number of data.

$$ST = \frac{PP}{UBP}$$

where

PP - number of verifiable data

UBP - total number of data

It is critically important to define criteria for the verification of data. It is strongly advised not to keep a fixed number but constantly update this list. For this research we started with kinds of self-evident criterias, such as the possibility of confirmation in transparent records, minutes, public reports, publicly verified documents, public registers, and transparent actions of the involved actors, such as direct tv interviews or any other similar direct source, five open sources that confirm information (today it is not applicable due to lots of web copying), basic observations, all depending on the data we want to verify. Criteria as not having numerus clausus need to be constantly updated. During the analysis we observed a few more. In the examples below we see that the author used a very general syntagm "Croatian citizens":

Data 5 Croatian citizens have started thinking about banks and the banking system that can hide many secrets - verifiable
Data 6 Croatian citizens have started thinking that the president could lie and be connected to crime if he has an advisor who can lie and be connected to crime - verifiable
Data 7 Croatian citizens began to think of other things as well - verifiable

We don't know their age, region in Croatia, education, even the number of people, so we could apply very simple sociological research, take an undefined sample of Croatian citizens (as here) and try to include 1,000 people so the research could be considered legitimate and verify/not verify this data. We can see that different data can be an inspiration for the indefinite number of criteria.

Specific time of the data publishing also turned out to be crucial in defining the criteria for the verification. The examples shown here are from the time when the Croatian Act on the right of access to information (OG 25/13, 85/15, 69/22) didn't exist in the legal system. As Article 3. of the Act states "The objective of this Act is to enable and ensure the exercise of the right of access to information and the reuse of information, as granted by the Constitution of the Republic of Croatia, to natural persons and legal entities, through openness and the publicity of the activity of public authority bodies." Therefore, from the moment the Act is valid in the country, we have many more verifiable data.

5. Conclusion

The criteria for data verification strongly influence the calculation of relevance that is defined as an unchangeable charateristic of the information as regards the publishing time, but as we stated before, we expect changes in the research results, even changes in the model itself, due to environment changes.

References

Act on the right of access to information, Official Gazette, 25/13, 85/15 and 69/22

Buzan B. (1983). People, States, and Fear: The National Security Problem in International Relations. Brighton, Sussex, UK: Wheatsheaf Books.

Floridi, L. (2009). Philosophical Conceptions of Information. In: Sommaruga G. (Ed.). Formal Theories of Information: From Shannon to Semantic Information Theory and General Concepts of Information (Lecture Notes in Computer Science, 5363) 13-53. New York, USA: Springer. https://doi.org/10.1007/978-3-642-00659-3_2

Floridi, L. (2011). The Philosophy of Information. Oxford, UK: Oxford University press.

Giannopoulos, G., Smith, H., Theocharidou, M. (2021). The Landscape of Hybrid Threats, A conceptual model. EUR 30585 EN, Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/44985

Gill, P., Phythian, M. (2018). Intelligence in Insecure world. Oxford, UK: Polity Press.

Jain A. K., Mao, J., Mohiuddin K.M. (1996). Artificial Neural Networks: A tutorial. Computer 29(3): 31-44. https://doi.org/10.1109/2.485891

Ma, L. (2012). Meanings of Information. The Assumptions and Research Consequences of Three Foundational LIS Theories. Journal of the American Society for Information Science And Technology 63(4): 716–723. https://doi.org/10.1002/asi.21711

McQuail, D. (2005). McQuail's Mass Communication Theory. Sage Publications Ltd., London.

McQuail, D. (2015). Political Communication Research in the Public Interest. In: Coleman, S., Moss, G., Parry, K. (Eds.) Can the Media Serve Democracy? 76-86. London, UK: Palgrave Macmillan. https://doi.org/10.1057/9781137467928_7

Oddie, G. Truthlikeness. In E. N. Zalta (Ed.), The Stanford encyclopaedia of philosophy (Winter 2016 ed.). Stanford University.

Saračević, T. (2006). Prilozi utemeljenju informacijske znanosti. Osijek, Hrvatska: Filozofski fakultet.

HISTORY FOR NON HISTORIANS: ANALYZING THE EFFECTIVENESS OF VARIOUS NATURAL LANGUAGE PROCESSING MODELS USED IN SOCIAL MEDIA

Jayashabari Shankar[1*]

[1] University of Utah, United States of America

*correspondence E-mail: jayashabari.shankar@gmail.com

Keywords: Natural Language Processing; social media; history; machine learning; popular history; historians; non historians

The rise of social media, particularly in the speed and quantity of data being transferred, has the potential to empower members of the public by facilitating their understanding of past historical concepts, whether to apply it to present day scenarios or to deepen their comprehension of contemporary problems and the root of their causes. Prior research on the impact of social media on "popular history" focuses on issues caused by both creators and users of "popular history" transmitted through social media such as simplification and dramatization of historical fact and the spread of misinformation. Here we undertake a detailed study of the algorithmic issues in spreading open data through social media, biases caused not necessarily by human social media users, but rather by the technological application itself. (More specifically, these technological "biases" can be described as human biases, simply exacerbated by the nature of the algorithms in the technological application. The applications considered here, however, do not have any inherent biases).

We focus particularly on social media algorithms broadly categorized under Natural Language Processing (NLP) machine learning because NLP deals with text data, the most common method of transmitting "popular history" to a general audience. Prior research mostly focuses on extracting information from social media sites, and then using NLP to process it, but here, we take a different approach. NLP is also used by social media sites to personalize one's experience. It is used in social media "search engines" (like Google Autocomplete) but mainly in personalizing one's social media "feed." The techniques we analyzed in the study (N gram, Bag of words, word embedding) are all used to generate new information given large bodies of text relating to a certain topic (Song and Croft, 1999). We took a "deep dive" into these techniques, considering the Python code used to write these algorithms, finding how bias in the bodies of information we feed the model can affect the work it produces. Ultimately, NLP is considered from a layperson/ nonhistorian's perspective instead of a historian's since we try to see the impact of these NLP techniques used by social media sites on the public's understanding of history.

To consider a more detailed methodology of the study, we started with an analysis of various NLP techniques used by various social media applications and features, such as N gram models, the "Bag of Words" model, and the word embedding approach (McCallum and Nigam, 1998). Using

probability distributions, we calculated the extent of algorithmic bias and then qualitatively determined the extent of the magnification of a human bias. Using our analysis of NLP models, we then predicted the impact that each NLP model had on the spread of human (whether creator or user) bias. The study finds that the word embedding algorithm works best for transmitting "popular history," as it both minimizes bias and also prevents the magnification of human errors (both simplification and dramatization of history).

Attached here are the more specific results from the study and details of each of the three techniques used (Gale and Church, 1994). The corpus, the data fed into the models, was a large database of a few hundred pages, on basic world history, obtained from a publicly available site (http://glhssocialstudies.weebly.com/world-history-textbook---pdf-copy.html).

For the N gram approach, or Markov model approach, these are the prompts generated:

George Washington is… (N = 4)
The world war was caused by… (N = 5)
what are some challenge right now from history.. (N = 8)
today Edison's discoveries are relevant… (N = 9)
Alexander the Great conquered these areas in the year of 300 BC and… (N = 12)

Notice how the model varies depending on the N we input. The higher the N, the more accurate it is, but it also means a higher likelihood of the model just copying the corpus (Franz and Brants, 2006). Additionally, the model has no sense of grammatical correctness and will often output incoherent or repetitive prompts, especially when N is small (Kneser and Hey, 1995).

In the Word Embedding approach, each word is represented by a vector, also called "word embeddings." Similar words have similar word embeddings, as their vectors will together have a higher dot product ("cosine similarity").

In the Bag of Words approach, the order of the words in the document does not matter because the model instead creates a "bag" of words without regards to structure and order. We analyze how unique a word is in a certain document or across all documents and the next words are found using associations (similarity scores).

The concept of open data is one that has percolated its way into public consciousness in the realm of history (Goodman, 2006). This study could improve the NLP models used by various social media sites, ensuring a better understanding of history for non-historians. The intersection between artificial intelligence algorithms in social media and the spread of "popular history" through social media is a valuable field to study and understand, especially as the world of history goes more and more technological. The public's understanding of history is also a crucial issue to consider, especially with the ubiquity of historical knowledge in the social media/ Internet age. Our study, by considering the issues facing the field of history through a technological lens, hopes to bridge the gap between public understanding and historians' understanding of history (Moosmann et al., 2007).

References

Franz, A., Brants, T. (2006). All our n-gram are belong to you. Available at https://ai.googleblog.com/2006/08/all-our-n-gram-are-belong-to-you.html

Gale, W. A., Church, K. W. (1994). What is wrong with adding one? InOostdijk N., de Haan, P. (Eds.). Corpus-Based Research into Language 189–198. Rodopi.

Goodman, J. (2006). A bit of progress in language modeling: Extended version. Technical Report MSR-TR-2001- 72, Machine Learning and Applied Statistics Group, Microsoft Research, Redmond, WA.

Kneser, R., Ney, H. (1995). Improved backing-off for Mgram language modeling. ICASSP, volume 1.

McCallum, A., Nigam, K.(1998). A comparison of event models for naive bayes text classification. In AAAI Workshop on Learning for Text Categorization, Madison, WI.

Moosmann, F., Triggs, B., Jurie, F. (2007). Fast discriminative visual codebooks using randomized clustering forests. In Scholkopf, B., Platt, J., Hoffman T. (Eds.). Advances in Neural Information Processing Systems 19, 985–992. MIT Press, Cambridge, MA, 2007.

Song, F., Croft, W. B. (1999). A General Language Model for Information Retrieval. In Proceedings of the 22nd ACM Conference on Research and Development in Information Retrieval (SIGIR'99) 4–9.

OPENNESS AND USABILITY OF BIG DATA OF THE PAST AND TIME MACHINE PROJECT – DIGITAL TRANSFORMATION OF THE EUROPEAN CULTURAL HERITAGE

Vlatka Lemić[1*], Thomas Aigner[1]

[1] University Office for Archives, University of Zagreb, Croatia
[2] Time Machine Organisation, Austria

*correspondence E-mail: vlemic@unizg.hr

Keywords: Time Machine project; Big data of the past; Cultural heritage; Digital transformation; Open Access

Cultural heritage sector in Europe consists of vast number of professional organisations, networks, infrastructures and platforms in various areas and domains, as reflection of Europe reach history and cultural diversity. In last decade they increasingly coordinate their work on building digital capacity in the cultural heritage sector by giving momentum to existing policies and developing of common practices and shared solutions, which is also supported by corresponding EU activities, such as for example, Expert Group on Digital Cultural Heritage and Europeana (DCHE).

Majority of European countries actively support digitisation and digital preservation of cultural heritage and provided access to the corresponding digitised resources, through various set of actions – ranging from policy initiatives and legislations like the Recommendation on the digitisation and online accessibility of cultural material and digital preservation (2011/711/EU) and the New European Agenda for Culture, programmes such as the European Year of Cultural Heritage 2018 or the European Framework for action of Cultural Heritage till building of the Europeana Europe's digital platform for cultural heritage which function as the European cultural hub. Majority of these activities are focused on development of digitisation technologies, digital preservation and innovative cultural projects that will make cultural heritage accessible to all and enhance its visibility.

Following these trends, in 2019 the Declaration of cooperation on advancing digitisation of cultural heritage was launched with aim to facilitate and promote joint work and better use state-of-the-art digital technologies in addressing risks that Europe's cultural heritage is facing, encompassing three pillars of action:

- a pan-European initiative for 3D digitisation of cultural heritage artefacts, monuments and sites;
- re-use of digitised cultural resources to foster citizen engagement, innovative use and spill-overs in other sectors;

- enhancing cross-sector and cross-border cooperation and capacity building in the sector of digitised cultural heritage ("EU Member States sign up to cooperate on digitising cultural heritage").

In times of global pandemic crisis caused by COVID-19 outbreak, importance of digital access to cultural heritage is additionally emphasized. European Heritage Alliance Manifesto „Cultural Heritage: a powerful catalyst for the future of Europe", launched in 2020 highlights seven incarnated ways in which cultural heritage can act as a catalyst for positive change, including digital transformation where Europe plays a leading role in digital cultural heritage and has the potential to forge ahead with new technologies such as artificial intelligence and machine learning based on humanistic and ethical principles". It also outlines collaboration the use of digital technology and expertise, innovation, narrowing digital gap between cultural institutions, as well as critical engagement in education and knowledge sharing (European Heritage Alliance, 2020).

EU-funded Time Machine Project, an ongoing Europe-wide initiative started in 2017 with the joint goal of digitising the entire European cultural heritage, builds on these foundations with idea of forming a unique alliance between the best European players in the humanities, sciences and technologies for implanting this goals. Time Machine, the largest and most ambitious project ever created at the intersection of culture heritage and information science, is built around the vision to develop the big data of the past – a widely distributed digital information system mapping the European social, cultural and geographical evolution across times by designing and implementing advanced new digitisation and artificial intelligence (AI) technologies. This large-scale digitisation and computing infrastructure, capable to map 5,000 years of European history, aims to revive historic data into (virtual) reality and convert them into a living social and economic resource that will start and influence various cultural, economic and social shifts (Time Machine Consortium, 2019a).

Time Machine plans to transform kilometres of archival fonds, abundant museum collections and various geo-historical data sets into a distributed digital information system by bringing together academic research teams, heritage institutions and industry, from major GLAM institutions to leading technology companies in the field of digitization and artificial intelligence. Such interdisciplinary operational environment is indispensable for developing the big data of the past that would launch a new era of open access to sources from history and culture field, as well as provide practical collaborative model for science and technology to actively contribute to safeguarding European identity and democratic values (Time Machine Consortium 2019b).

Time Machine is organised through four basic pillars: Science and Technology, Time Machine Operation, Exploitation Avenues and Outreach and Innovation while project infrastructure is implemented through Time Machine Organization (TMO).

The Time Machine ecosystem includes academic, research and cultural institutions, as well as large businesses and innovative small and medium-sized enterprises, government bodies and associations of the civil society involved in cultural heritage which make it the widest initiative for digitizing European cultural heritage, created at the intersection of culture heritage and information sciences. This collaborative network target specific objectives:

- addressing the scientific and technological challenges in artificial intelligence (AI), robotics and ICT for social interaction in order to develop the big data of the past, while boosting these key-enabling technologies in Europe,
- building the infrastructure for digitisation, processing and simulation that will support a sustainable management and operational model (the "TM franchises" in the form of local Time Machines), as well as create the basis for and engagement with communities (citizens, scientists, innovators) participating in the development and use of Time Machine,
- creating innovation platforms in promising application areas, by bringing together developers and users for the exploitation of scientific and technological achievements, and therefore leveraging the cultural, societal and economic impact of Time Machine,
- fostering favorable framework conditions for the outreach to all critical target groups, and for guiding and facilitating the uptake of research results produced in the course of the Time Machine initiative (Time Machine Consortium 2019b).

TMO is formally established in 2019 and operates as sustainable joint platform for future development and research in technology, science and cultural heritage and their positioning in Horizon Europe programs, as well as framework for strategic planning and lobbying activities. Today this global initiative gathers thousands member organisations, thus establishing the largest European network of academic, research and cultural institutions and the IT sector, that allows cross-sector communication and partnership structure capable to provide required global collaboration and resources for mass-scale digitization and shared computing infrastructure, as well as explore potentials related to research, innovation, technological development and cultural heritage. (see Time Machine Consortium, 2019b). TMO is headquartered in Vienna, with additional offices in Lausanne, Amsterdam and Budapest, engaged in building of network and infrastructure that implements the goals and activities of the Time Machine road maps(servers, databases, platforms, etc.), supporting local TM projects, organizing national TM days, TMO conferences, developing new services and projects, etc.

References

European Heritage Alliance. (2020). European Heritage Alience Manifesto – Cultural heritage: a powerful catalist for the future of the Europe. Available at https://www.europanostra.org/wp-content/uploads/2020/05/20200509_EUROPE-DAY-MANIFESTO.pdf. Accessed on August 6, 2020.

European Commission. (2019). EU Member States sign up to cooperate on digitising cultural heritage. Available at https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-digitising-cultural-heritage. Accessed on August 7, 2020.

Time Machine Consortium. (2019a). Time Machine roadmaps mapping the future. Available at https://www.timemachine.eu/time-machine-roadmaps-mapping-the-future/. Accessed on June 10, 2020.

Time Machine Consortium. (2019b). Time Machine Manifesto: Big Data of the Past fot the Future of Europe. Available at https://www.timemachine.eu/wp-content/uploads/2019/06/Time-Machine-Manifesto.pdf. Accessed on June 6, 2020.

SESSION VIII

POSTER SESSION

OPEN, USEFUL AND REUSABLE DATA (OURData) POLICY IMPACT ON STARTUP INNOVATION

Anton-Jan Klasinc[1*]

[1] Croatian Institute of Public Administration, University of Zagreb, Croatia

*correspondence E-mail: ajklasinc@gmail.com

Keywords: open data; startup innovation; open government; good governance; data policy

The OURdata index benchmarks the design and implementation of open data policies at the central level and stresses the sustained political and policy relevance of this areas of work for OECD member and partner countries and beyond,. Open Government data creates social and business values and are a core component of strategies aimed at strengthening good governance (OECD, 2022), and they have a significant potential for business applications (WB, 2021) However, it may be possible that some obstacles are preventing the full benefit of open government data in terms of innovation, and in this paper we aim to study these obstacles and, possibly, catalysts for better usage of open data in startup innovation.

Open data index is based on three pillars: open data availability, accessibility and government support for data reuse. Six dimensions of Digital Government are government as a platform, open by default, data-driven public sector, digital by design, user driven and proactive. The index is based on the analytical framework for open data policies, namely OECD Working Paper Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives (Ubaldi, 2013) which led to current OECD index in 2917 and 2019. It can be shown that just publishing open data does not automatically yield economic positive effects, which was also evident in Croatia, but accessibility and also networking is equally important. Also, Croatian startups lack an actual eco-system in areas such as AI, so a lot of risk is involved in making such business ideas work. EU projects and funds may be of some help, but not if the competitions are organized on the basis of the 'fastest finger', i.e., only the speed of submitting an application is a criterion for winning the tender,. However, recently we have seen many successful Croatian startups appear, so they must have been able to overcome the aforementioned problems. The question remains how much open data contributed to the success of Croatian startups. Business areas that appear as alluring are tourism, food, construction and health, but also transport and sustainability. In the area of transport and sustainability open government data may be of some importance. For investment into construction, availability of government data concerning ownership of properties may also play an important role for new startups.

In this paper we analyze how Croatian open data policies have impacted startups and we try to find concrete examples of startups that have benefited from those policies, we enumerate and analyze those policies and compare them to similar policies in EU countries, in order to draw conclusions about their efficiency and effectiveness.

Open government data promotes transparency, accountability and value creation. This last aspect is of particular interest when researching startups and innovation, but transparency and accountability are equally important for a successful and functional open data ecosystem. This ecosystem in Croatia is determined by e-government strategy and decision to make data open by design. However, OECD countries have become aware that formal government open data requirements are essential, but not sufficient, which may even give rise to "open washing" and the accessibility may be even more important than availability of open data, as availability does not always result in implementation and availability is more important than quantity of open data (Ubaldi, 2013), which may turn to be a problem to achieve in case of Croatia and other countries that are less successful than the leaders in open data, namely Korea, Canada, France, Ireland, Japan, as political culture and admninistrative structure (multi-level) may be reason for slowing down, and governments should also monitor and promote re-use and monitor economic and social impact of open data in order to determine better policies. (OECD Open Data Index, 2022).

It has been determined that Open data play a role in startup innovation and entrepreneurial activity (Hubert et al., 2022). In our case we are interested in value creation in the economic sense, and for that we must determine the quality of Croatian startup ecosystem, and see if it can also be linked to open data policies. Luckily, there is such an index, and on this index the leading countries are USA, UK and Israel. Croatia has recently dropped 9 places and is now 45th and 9th in the SEE region (Global Startup Ecosystem Index, 2022). A comparable country is Slovenia, which is similarly positioned on the 47th and 11th place and may serve as a benchmark, although most startups in Slovenia are in foodtech and transportation, whereas in Croatia it is leisure (hotels etc.), green tech and energy, as well as IoT (sensors) (GSEI, StartubBlink). Data accessibility is crucial for startup companies, as they have to collect good data first in order to implement data-driven approach. It becomes clear that one of main startups for which open data are crucial is Rimac Automobiles, ltd., because it makes electric cars that depend on geospatial data, which are provided by government agencies, notably Shared Service Center (SSC), in making, and the Croatian Ministry of Physical Planning, Construction and State Assets.. In fact, it can be seen that Croatian e-government strategy has played an important role in leading up to this development in the past 10 years (Vračić, 2014), and those startups in Croatia did not appear out of the blue, as it can result in better entrepreneurial strategies (Kitsios et al., 2018). In fact, the type of country determines competitive advantage (Porter, 1998) and in this case Croatian strategies assisted by EU funds have played into this hand by enabling the right kind of development (Saberi et al., 2019), leading to the innovation in electric cars, sports cars in nature (which is naturally linked to leisure as a specific feature of Croatia, another is also roads, love for sports, etc.) and in the future also autonomous cars, SAE level 5 (without drivers). Of course, many startups did in fact come out of the blue or by the efforts of their makers, such as Bellabeat or Infobip, but the overall climate for startups has improved due to the development of e-government and abolishing of outdated practices (seals, etc.), while introducing e-citizens and other modern features of information society. The next step in the SSC project that was started in December 2019, would be to actually monitor economic impact in the Ministry of Economy in order to determine better ways to support businesses. The role of the EU as a funding and regulatory force is also of primary importance, although the non EU countries are leaders in startup innovation (US, UK and Israel), but France and Germany are not too far behind.

In this paper we were able to prove that open government data can indeed play an important role in startup innovation, and a rather straightforward example of Rimac Automobiles ltd. can serve as a good epitome of such a link, implying causality. However, there is much left to be desired in terms of using open government data for startup innovation, as well in terms of accessibility of such data, which are not exploited to the full extent or presented in a suitable manner for startups to be able to profit from them or get a great business idea. Many Croatian startups are in fact directed towards households and citizens, as consultancies that explain how they can better analyse their own data in order to improve their daily functioning, e.g. in terms of energy saving, driving around and parking more efficiently, saving unused food, disposing garbage, etc.

References

Gertmo, E., Hormez, A. (2019). How to find potential business ideas using Open Data – case study on Swedish Open Data Portal and Road Safety Data in UK., Research assignment fo Exchange Students.

Global Startup Ecosystem Indeks. (2022). StartupBlink (Report)

Huber, F., Ponce, A., Rentocchini F., Wainwright, T. (2022). The wealth of (Open Data) Nations? Open government data, country-level institutions and entrepreneurial activity 992-1023.

Kitsios, F.C., Kamarioutou, M. (2018). Open data and high-tech startups: Towards nascent entrepreneurship strategies. In Khosrow-Pour, M. (Ed.). Encyclopedia of Information Science and Technology, Fourth Edition, IGI Global Publishing 3032-3041.

OECD Open Data Indeks. (2022). OECD, Paris.

Porter, M.E. (1998). Competitive advantage of nations, Free Press.

Porter, M.E. (1998). Sustaining competitive advantage, Creating and sustaining superior performance. Free Press.

Saberi, M. W., Hamden, A. (2019). The Moderating Role of Governmental Support in the Relationship between Entrepreneurship and Economic Growth: A Study on the GCC Countries. Journal of Entrepreneurship in Emerging Economies 11(2): 200-216. https://doi.org/10.1108/JEEE-10-2017-0072

Somaya, D., Teece, D. (2021). Patents, licensing, and entrepreneurship: Effeftuating innovation in multi-invention contexts. In Shashinski, Č E., Strom, R.J., Baumol, W.J. Entrepreneurship, innovation and the growth mechanism of the free enterprise economise, US; 185-212, Princeton: Princeton University Press.

Temiz, S., Holgerssson, M., Björkdahlb, J., Wallin, M.W. (2022). Open data: Lost opportunity or unrealized potential. Technovation 114: 102535. https://doi.org/10.1016/j.technovation.2022.102535

Tinholt, D. (2013). The Open Data Economy, Unlocking Economic Value by Opening Governmetn and Public Data. Capgemini Consulting.

Ubaldi, B. (2013). OECD Working Paper Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives, OECD, Paris.

Vračić, T. (2014). Otvoreni podaci: Pristup informacijama u Republici Hrvatskoj: novi izazovi, nova rješenja.

World Bank Group (2021). The Huge Potential of Open Data for Business Applications, Washington DC, US: World Bank Group.

ACCESSING OPEN DATA ON NFT

Xinyi Zhu[1*]

[1] School of Social Sciences, Faculty of Humanities, The University of Manchester, UK

*correspondence E-mail: xinyi.zhu-9@postgrad.manchester.ac.uk

Keywords: NFT; Digital Society; Open Data; Blockchain; Walkthrough Method

As social media has played an essential role in expressing individual ideas, connecting, communicating, and interacting with one another on the digital platform, it has also provided a novel set of data to understand our daily life (Kakulapati, 2021). Non-fungible token (NFT) is an extension of digital token transactions that emerged in 2021. It has also inherited a much more advanced set of data and provides an alternative for information and data communication. As NFT differs from fungible tokens, it is the presentation of the digital and unique transactions of tokens. Based on the blockchain system, NFT is intrinsically the demonstration of decentralised system applied in the transaction of digital work. Each property can be traded freely with the customized value that is set by the NFT creator (Wang et al., 2021). As the value is being set and generated through the smart contract, here the smart contract refers to the adapted contracts that are run by programs that are originally set by individuals, this means that every NFT transaction has complete freedom of setting its own contract conditions, with external interference. NFT transaction is also a transparent process (Wang et al., 2021). While the NFT transaction has been finished, there is a smart contract that has been generated through machines automatically and it is not reversible or editable. The whole transaction process and related information are all documented fully by the machine. Due to the existing transparent framework and the decentralised system, the open data on NFT transactions and predictions are mushrooming ever since 2021.

This research intends to examine the platforms for the current open data access on NFT works and analyse the potential social impacts on the understanding of NFT. The research follows the new materialism philosophy as the main guidance, applying the Actor-network Theory (ANT) to connect technology and social life (Latour, 2007). The aim is to combine the digital technical examination of NFT trading platforms and examine how the open data on the trading site has impacted the social aspect, which refers to the general users' experience and perceptions of NFT. The research will firstly conduct an overview explanation of the NFT concepts, open data definition, and the implication of social media open data on digital platforms, and how the transparent framework allows more data to be retrieved and tracked online to be accessed by the public. Furthermore, the research uses the Walkthrough method to thoroughly investigate the step-by-step process of the open data platform in NFT. It is the digital observation of the NFT user experience from registration, profile building, daily interactions, and any other changes of status (Light et al., 2018). The Walkthrough method is also suitable for social media platform examinations, in order to understand how specific functions and technical frameworks have an impact on the users' social life. To complement the observation data, I explore NFT from various

social perspectives, including how the open data has access to the encrypted information, how the data has been displayed, and how it effectively impacts the transaction of NFT. The conclusion is drawn on the potential social impacts on users' daily experience of the NFT applications and identifies the significance of open data access for NFT work, in terms of the NFT transaction. As NFT is a relatively underdeveloped research topic, this will be the first research on the interdisciplinary research of data science, NFT technologies, and social understandings. The research is aiming to provide references for further research on accessing open data through various digital platforms and its social impacts.

References

Kakulapati, V. (2021). Knowledge Extraction from Open Data Repository. In Open Data. IntechOpen. https://doi.org/10.5772/intechopen.100234

Light, B., Burgess, J., Duguay, S. (2018). The walkthrough method: An approach to the study of apps. New Media & Society 20(3): 881–900. https://doi.org/10.1177/1461444816675438

Wang, Q., Li, R., Wang, Q., Chen, S. (2021). Non-Fungible Token (NFT): Overview, Evaluation, Opportunities and Challenges.
Available at https://www.semanticscholar.org/paper/ad5c04b15e20dccfd3907f9cc354962cb77848a7

Table of contents