sciendo

Categorisation of Open Government Data Literature

Aljaž Ferencek, Mirjana Kljajić Borštnar
Faculty of Organizational Sciences, University of Maribor, Slovenia
Ajda Pretnar Žagar
Faculty of Computer and Information Science, University of Ljubljana, Slovenia

Abstract

Background: Due to the emerging global interest in Open Government Data, research papers on various topics in this area have increased. **Objectives:** This paper aims to categorise Open government data research. **Methods/Approach:** A literature review was conducted to provide a complete overview and classification of open government data research. Hierarchical clustering, a cluster analysis method, was used, and a hierarchy of clusters on selected data sets emerged. **Results:** The results of this study suggest that there are two distinct clusters of research, which either focus on government perspectives and policies on OGD, initiatives, and portals or focus on regional studies, adoption of OGD, platforms, and barriers to implementation. Further findings suggest that research gaps could be segmented into many thematic areas, focusing on success factors, best practices, the impact of open government data, barriers/challenges in implementing open government data, etc. **Conclusions:** The extension of the paper, which was first presented at the Entrenova conference, provides a comprehensive overview of research to date on the implementation of OGD and points out that this topic has already received research attention, which focuses on specific segments of the phenomenon and signifies in which direction new research should be made.

Keywords: Open government data; Open government data research; Hierarchical clustering; OGD classification; OGD literature overview
JEL classification:
Paper type: Literature review

Received: Oct 27 2022
Accepted: May 8 2022

Acknowledgements: This research was supported by the Slovenian Research Agency: Program No. P5-0018—Decision Support Systems in Digital Business.

Citation: Ferencek, A., Kljajić Borštnar. M., Pretnar Žagar. A. (2022), "Categorisation of Open Government Data Literature", Business Systems Research, Vol. 13 No. 1, pp.66-83.
DOI: 10.2478/bsrj-2022-0005

Introduction

Organisations must adapt to the environment in which they do business. Organisational development, an activity that utilises research findings further to improve organisations' functioning (Buchanan, 1972), has existed in the organisational ecosystem for quite some time. Therefore, researching organisational changes and development has offered different explanations, research studies and theoretical models. From the point of understanding organisational changes as described by Lewin's three-stage model of changes (Lewin, 1947), establishing change process theories (Van De Ven et al., 1995), improving them with Peters and Watermans' Culture-Excellence approach (1982) or furthermore, applying another dimension of inter-relatedness of individuals in the organisation by Pettigrew (1973). Although different models may not explain every real-world situation, they provide the grounds for designing, planning, and implementing changes (Maxwell et al., 2015). Several authors improved theories and models in the past years to adapt them to the new business environment. Relying on the four change process theories by Van De Ven et al. (1995), authors (Crossan et al., 2009; Weick et al., 1999; Tsoukas et al., 2002) provided another dimension of taking the macroscopic and microscopic level of analysis into account, making microscopic changes to seem less routined and much more agile to changes. Another observation by Weick et al. (1999) states that organisations tend to have small groups of people open to innovations, thus making them the main drivers for adjusting to the always-changing business environment. The latter is also supported by Damanpours et al. (2008) research on the importance of personal characteristics in adopting innovation in public organisations. The challenge of accepting continuous change throughout an organisation to benefit more generally remains.

A generalisation of different theories or models on both private and public sector, routine and bureaucracy seem to be linked to public organisations much more often, making a general belief that public and private management are fundamentally dissimilar (Boyne, 2002). Nevertheless, Boyne's (2002) findings indicate that despite several differences, such as human resource practices, ethical issues and decision processes, there is no support to prove that both managements are different. As private organisations tend to provide services or products to consumers, local governments are very important in delivering services to the public (Walker et al., 2013). Since local governments provide public services on which citizens rely, they need to be open to innovation and organisational changes, likely to continue to some extent (Elliot, 2020). The government sector must ensure the effectiveness of new technological tools, such as e-government, because it requires support and trust from the citizens, who have to be reassured of their financial input - taxes (Nam, 2012) and can, by innovating, contribute to effectiveness in the production of services which increases business performance quality (Marčeta et al., 2020).

Since this paper aims to examine and classify current research on Open Government Data (OGD) in the public sector based on a literature review, we should first examine practices of relatively similar yet strongly related research fields; e-government. E-government can be defined as a construct that uses information technology (IT) to improve communication between governments and other actors (Sprecher, 2000). Because e-government was established after the private sector adopted e-business and e-commerce (Moon, 2002), this pattern is expected to be followed in data gathering and sharing. West's (2004) research on e-government development shows four primary stages of transformation: the billboard stage, the partial-service-delivery stage, the portal stage and interactive democracy with public outreach. We could assume that today's open government data development stage

is in the third or the portal stage of transformation since there are many different OGD portals worldwide without being used in a broader context (Lourenço, 2015). In their research, Tolbert et al. (2006) found out that with the frequency of e-government services, usage does not correlate with the enjoyment of greater trust among citizens on a federal government level. This gap could be filled by opening government data to the broader public since it represents the considerable potential for improvement and transparency, thus making it substantial for e-government services (Bertot et al., 2014). Nevertheless, we should not perceive those two terms as equal since the main difference is that open government is represented as one of the main parts of e-government, differentiating with its ultimate goals and promises in politically based agenda (Kassen, 2013; Harrison et al., 2011).

The problem addressed in this paper is that OGD has received a lot of research attention in the past ten years since open data is a prerequisite for companies and public administration to innovate (Jetzek et al., 2013; Vetrò et al., 2016; Welle Donker et al., 2017) successfully. Nevertheless, the research has been widely fragmented into many different areas. There is no extensive summary of the current state of research, future research recommendations, and categorisation of research, which could lead to new ideas and knowledge discovery in this field. For this purpose, we conducted a systematic review of the literature to determine what authors were focusing on, the latest research gaps and in what areas of interest we can classify current research on OGD.

Methodology

In this section, the review scope and selection of papers are presented. We used a Systematic literature review (SLR), the primary purpose of which is to either identify gaps in current research or to point out possible future research areas (Kitchenham, 2004; Webster et al., 2002). Webster et al. (2002) described that reviewing existing literature is essential for further academic research and knowledge advancement. To examine the current state of research, we created a meta-review or overview of existing literature reviews on OGD implementation and categorised and examined their findings. The papers included were peer-reviewed conference and journal articles from Scopus and Web of Science databases. Our search targets were existing literature reviews in open data implementation using "open government data" AND "literature review" in the article title, abstract, and keywords.

Since investigating the implementation of OGD manifests in many different forms of research, our primary criteria for selecting the papers were:

1. We only included existing literature reviews on OGD implementation.
2. Literature reviews were done on at least 25 referencing articles.
3. Each literature review had to examine OGD implementation on either international or country level, and their objective either had to: (i) Examine the implementation of OGD; (ii) Analyse the dissemination of OGD and (iii) Compare and break down OGD initiatives, their policy-making, along with barriers and their adoption.

Examination of the papers showed that the topic "OGD implementation" can be found under the following phrases: OGD initiatives, barriers or challenges of OGD, use of OGD, dimensions of OGD, adoption of OGD, barriers of OGD, OGD policy, development of OGD, conditions of OGD implementation, OGD perspectives, OGD initiatives, OGD ecosystems, and OGD dissemination.

Twenty-two initial papers for this literature review are presented in Table 1, where each article is defined by its author, year of publishing and corresponding title.

Table 1
Selected previous research.

Nr.	Author	Year	Title	Cluster
1	Ali Hassan, M. & Twinomurinzi, H.	2018	A Systematic Literature Review of Open Government Data Research: Challenges, Opportunities and Gaps	B
2	Attard, J., Orlandi, F., Scerri, S. & Auer, S.	2015	A Systematic Review of Open Government Data Initiatives	A
3	Chatfield, A., Reddick, C. & Al-Zubaidi, W.	2015	Capability Challenges in Transforming Government through Open and Big Data: Tales of Two Cities	A
4	Crusoe J. & Melin, U.	2018	Investigating Open Government Data Barriers: A Literature Review and Conceptualization	B
5	Csáki, C. & Kő, A.	2018	Open Data Research Challenges in the EU	A
6	De Oliveira, E. F. & Silveira, M. S.	2018	Open Government Data in Brazil: A Systematic Review of its Uses and Issues	A
7	Gil-Garcia, J., Gasco, M. & Pardo, T.	2020	Beyond Transparency, Participation, and Collaboration? A Reflection on the Dimensions of Open Government	A
8	Haini, S., Rahim, N. &	2019	Adoption of Open Government Data in Local Government Context: Conceptual Model Development	A
9	Hossain, M., Dwivedi, Y. & Rana, N.	2015	State-of-the-art in Open Data Research: Insights from Existing Literature and a Research Agenda	A
10	Huang, R., Lai, T. & Zhou, L.	2017	Proposing a Framework of Barriers to Opening Government Data in China: A Critical Literature Review	B
11	Ingrams, A.	2016	An Analytic Framework for Open Government Policy Design Processes	A
12	Kalampokis, E., Tambouris, E. & Tarabanis, K.	2011	A Classification Scheme for Open Government Data: Towards Linking Decentralised Data	B
13	Katbi, A.-K. & Al-Ammary, J.	2019	Open Government Data in the Kingdom of Bahrain: Towards an Effective Implementation Framework	B
14	Nugroho, R, Zuiderwijk, A. Janssen, M. & de Jong, M.	2015	A Comparison of National Open Data Policies: Lessons Learned	A
15	Roa, H., Loza-Aguirre, E. & Flores, P.	2019	A Survey on the Problems Affecting the Development of Open Government Data Initiatives	A
16	Safarov, I., Meijer, A. & Grimmelikhuijsen, S.	2017	The utilisation of Open Government Data: A Systematic Literature Review of Types, Conditions, Effects and Users	B
17	Saxena, S.	2018	Summarising the Decadal Literature in Open Government Data (OGD) Research: A Systematic Review	B
18	Susha, I., Johannesson, P. & Juell-Skielse, G.	2016	Open Data Research in the Nordic Region: Towards a Scandinavian Approach?	A
19	Wirtz, B., Piehler, R., Thomas, M. & Daiser, P.	2015	Resistance of Public Personnel to Open Government: A Cognitive Theory View of	A

			Implementation Barriers Towards Open Government Data	
20	Yuan, Q.	2019	Co-production of Public Service and Information Technology: A Literature Review	**A**
21	Zuiderwijk, A. & Hinnant, C.	2019	Open Data Policy-making: A Review of the State-of-the-art and an Emerging Research Agenda	**A**
22	Zuiderwijk, A., Janssen, M. & Davis, C.	2014	Innovation with Open Data: Essential Elements of Open Data Ecosystems	**A**

Source: Own

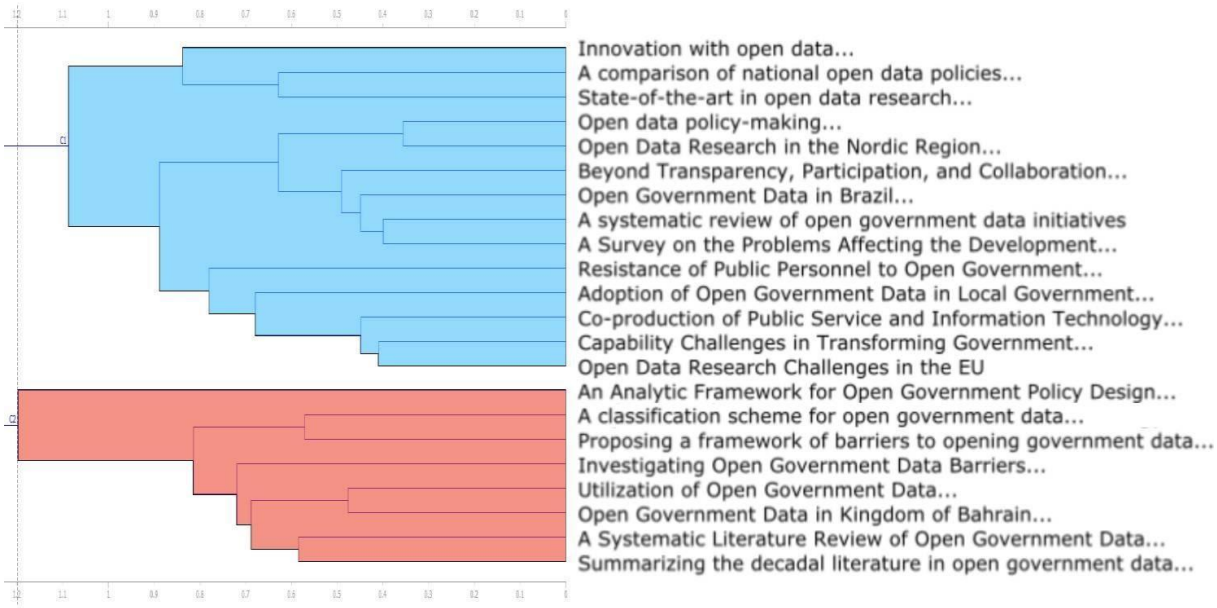
Since selected papers are different in geographical outreach, year of publication and approach to researching OGD initiatives, we seek to find out how existing literature reviews are similar or how they differ from one another. We performed hierarchical clustering on our corpus to determine which papers are similar. We used an open-sourced framework called Orange that is component-based and suitable for machine learning and data mining operations (Demšar et al., 2004). The hierarchical clustering results are presented in the last column of Table 1, which indicates the cluster in which each article was classified.

To perform hierarchical clustering in Orange, we had to preprocess our data. Since Orange can accept many different file formats as input, we decided to prepare our data in Tab-separated values (.tab). Preparing such a file was to export every research paper from the original PDF to plain text and then copy its content to a plain text editor and structure the features with a tabulator. We prepared four features; title, abstract, references and article content. Content that we copied from the selected articles did not include headers of the publication and other formal data such as copyrights, correspondence addresses and organisations. We also did not include any graphical material in our corpus due to the inability to transform it into plain text. We also decided to cluster abstracts rather than the content since abstracts provide a more precise definition of an article's research scope.

Another step in preprocessing the data was eliminating all the stop words using the Preprocess Text widget in Orange. Removing stop words is an often-used procedure to reduce the noise of textual data by using pre-compiled lists of stop words such as "in", "and" "the", etc. (Saif et al., 2014). To group the data and display it accordingly, we also need to measure the elements and their relative distances, which helps us decide which elements belong to a specific group (Murtagh et al., 2011). The computation can be performed with similarity/dissimilarity measurement or a more complex distance metric. Since one of our limitations is the distances provided in Orange, we used the metric recommended in the literature. The cosine metric was applied to our corpus since it is one of the most used and well-resulting distance measures in text classification problems (Al-Anzi et al., 2017; Lin et al., 2014). We used Ward linkage for cluster distance.

The result of the hierarchical clustering process is presented with a dendrogram in Figure 1, where we can distinguish between two clusters of research papers (cluster A (blue) and cluster B (red)).

Figure 1
Initial clusters of research papers



Source: Author's illustration

After the data preparation and analysis phase, we grouped the articles in two clusters, A and B. Groups of articles are presented in the fifth column in Table 1.

We compiled another data set to validate the initial clusters and put them in the broader research context. We retrieved the data from the ScienceDirect service with the keywords "open government data, OGD initiatives, barriers of OGD, challenges of OGD, use of OGD, dimensions of OGD, adoption of OGD, implementation of OGD, OGD policy, development of OGD, OGD perspectives, OGD initiatives, OGD ecosystems, or OGD dissemination" appearing anywhere in the title, content or abstract of the paper. We used ScienceDirect because it provides programmatic access to structured content in a .xml format. The search resulted in 54 additional papers, one of which appeared in the original cluster and was discarded.

By segregating our corpus, we intended to distinguish articles from one another and provide a classification scheme that could help us identify similarities in the discovered clusters. The second corpus served as a validation of the findings.

We used the same preprocessing with added lemmatisation to enable a more precise topic comparison. Next, we mapped the following data set to the same document space as the original corpus, which enabled a more precise definition of found clusters and elaboration of identified topics. Document maps were computed with t-SNE (van der Maaten et al., 2008), an embedding technique that optimises the instance neighbourhood. Finally, we used the YAKE keyword extraction algorithm to explain the clusters and local neighbourhoods in t-SNE (Campos et al., 2020). We manually extracted words or concepts common to instance locality.

The following two chapters present the literature review and cluster validation with the second corpus.

Results

Findings from the literature review are presented as short summaries of articles from each cluster, followed by cluster validation with the second dataset in the section "Cluster validation", and finally, a conclusion of this research paper and recommendations for future research in the section "Discussion and conclusions".

Cluster A

Ali Hassan et al. (2018) first research in cluster A captured the status of OGD research from 2012 – 2018. The primary objective was to identify opportunities, obstacles, and gaps to determine what has been researched in OGD. Their findings suggest that most research has been done in developed countries and that there is a lack of research on OGD for business benefit. Regarding the type of research, some theoretical papers explained the OGD initiative, yet many more Case Studies and Surveys were conducted in this field. Considering the practical implications of OGD, most research focused on proposing different frameworks and less on developing platforms. The main obstacles are related to the security and privacy of opening government data, government data formats, and the legal implications of opening up the data, such as copyright and licensing.

Addressing different types of OGD research, Haini et al. (2019) researched the shortcomings of empirical studies on adopting OGD in the local government context. Influencing factors of OGD adoption were identified, analysed, and reviewed by five selected experts from local governments. Experts validated and individually ranked sixteen influential factors as relevant, but only eleven were selected as strongly influential. Among those, top management was perceived as the most influential factor. As a result of this study, the authors proposed a conceptual model of OGD adoption for local government, classifying influential factors in three independent variables - technological, organisational and environmental.

Another comprehensive research on OGD initiatives assessment was made by Attard et al. (2015). The study's main research question addressed existing approaches for publishing and consuming government data. Research provides an explanation and types of OGD implementation, defines motivation for governments to open their data based on democracy and economics and introduces the OGD life cycle with the conclusion that several open data life cycles already exist. However, none of them is adjusted to the needs of OGD. Since government initiatives are usually evaluated according to their compliance with the law and not by their usefulness, authors also debate and classify different assessment frameworks and initiative evaluations. Finally, an overview of success factors has been made, and aspects of publishing and consuming OGD were investigated. Challenges or barriers to OGD implementation were classified based on the nature of the challenge into five major groups; technical, policy/legal, economic/financial, organisational, and cultural. With all the aspects considered, Attard et al. (2015) research managed to become one of the highest-cited research papers on the subject.

Hossain et al. (2015) also reviewed current research on OGD and presented findings for 11 types of analyses. Authors first classified the context of research on either government or the public, then they also classified different ways to explain open data movement. Another field of investigation was the level of analysis, where the aim was to distinguish whether analysis was made on individual cognition of open data, organisational/societal level or if analysis dealt with open data on an abstract level. Regarding the research methods, the authors' findings indicate that the majority, or almost fifty percent of the studies, applied a qualitative approach, followed by conceptual and quantitative papers. Drivers of OGD implementation are primarily

perceived in terms of political leadership, institutional pressure, or technology development, whereas barriers are primarily present at institutional, legal, economic, and technological levels. The latter is recognised as one of the most important due to its frequency of appearance in various research papers.

Zuiderwijk et al. (2014) overviewed the most important elements of open data ecosystems for simple publication and use of open data. Twenty essential elements of open data ecosystems were identified where the focus was either on the data producer or data user. Four essential elements were also identified: releasing and publishing data on the internet, searching and viewing the data, analysing, enriching, and visualising the data, and finally, discussing the data by providing sufficient feedback to data providers and other stakeholders. In addition to the four key elements of open data ecosystems, three elements were identified to integrate the other twenty ecosystem elements as a whole, and those are; different pathways on how to use open data, a quality management system that provides dataset satisfaction for users and finally the use of metadata that would offer interoperability and connection of the elements.

Gil-Garcia et al. (2020) researched open government dimensions to understand the term in a socio-technical context. Articles that were selected included a mix of conceptual or methodological research methods such as literature reviews, case studies, comparative analysis, etc., indicating the phenomena' importance. Five dimensions of open government were identified that provide different concepts related to transparency, participation, information availability, collaboration, and information technologies. The same concepts are sometimes identified as dimensions of open government or drivers of open government initiatives. The information availability dimension can be perceived as the most important since it discusses the intentions of the governments to make information publicly available and is, along with transparency, the most debated topic among researchers.

Csáki et al. (2018) focused on the relevant studies to review the research area of OGD in a more specific context of the European Union (EU). Open data and related challenges are debated mainly from a policy or technical point of view. Therefore, the authors conclude that a holistic view of OGD is missing. Authors further classify research areas of open data into nineteen different topics. The most often researched are policy and regulation, organisation and management, participants and roles (stakeholders), and technology and infrastructure.

Narrowing the scope of open data research in the EU, Susha et al. (2016) addressed the problem of the non-existing systematic overview of open data research in Nordic countries. Forty-four research papers were included in this literature review, examining research perspectives and topics. The main topics of Nordic open data research include open innovation, open data adoption, evaluation, benefits and barriers, etc.. In contrast, perspectives of OGD were adopted from Zuiderwijk et al. (2014). The authors' findings indicate that Nordic countries' social and economic research perspectives dominate. Innovation-related topics such as open innovation, open data entrepreneurship, service innovation, and innovation contests were the most popular research topics.

With a more specific focus on the case of Brazil, De Oliveira et al. (2018) intended to seek out what initiatives are being conducted there, how OGD is being used, and what prevents its effective use through presented challenges. Their findings indicate that transparency is the most popular aspect of OGD research in Brazil. In contrast, other research topics include the benefits of OGD in crime investigation, election data, biodiversity data sharing, and education. Finally, the authors conclude that the

challenges of OGD are considering the quality and format of the datasets, thus making the technical aspect of OGD important once more.

The following two papers focus more on barriers or the main problems affecting open government data initiatives. Starting with Roa et al. (2019), the authors used a literature review of the past sixty-nine OGD initiatives from 2012-2018. They identified six problem categories: citizen participation, data quality, economic and financial, organisational, policy and legal-related, and finally, technical. Out of all problem groups, data quality and policy-related problems are most commonly reported, whereas citizen-related and economic problems are least represented. Authors also discovered that most reported problems occur during the post-adoption stage, whereas during the adoption of open data, the report of problems of any group is relatively low. Policy and legal problems are most common in the pre-adoption stage.

Focusing on perceived barriers to introducing OGD, Wirtz et al. (2015) summarised factors that prevent successful implementation of open government data or resistance, as the authors define it, in a model based on cognitive theory and literature review. Since cognitive theory distinguishes perceived barriers on individual internal, organisational, and external, the final research model is also classified as such. Perceived legal barriers were classified as external, whereas internal organisational barriers consist of perceived bureaucratic decision culture, organisational transparency, and hierarchical barriers. Finally, a perceived risk related to administrative employees' attitudes is the only internal barrier. After the classification, a survey was conducted on thirty-five public authorities, resulting in 265 responses. The results of factor analysis indicate that the perceived risk-related attitude of the administrative employees turned out to be the most influential in connection with the open government data resistance, primarily due to the protective mechanism of individuals.

One of the often-debated subjects regarding OGD implementation barriers dealt with regulatory issues surrounding the release of government data. Nugroho et al. (2015) compared national open data policies from five countries, focusing on policy-making aspects. The comparison revealed that open data should become the default option instead of governments waiting for various requests to open data. The authors identified two waves of policy-making. The first wave proposes adapting the legal framework that would regulate and stimulate the continuous release of data. The second wave of policy-making focuses on how data providers and users interact to stimulate OGD usage. The authors suggested that there might be a third wave emerging where public forums and other participatory medians are created, enabling data users to provide feedback on quality and usability.

Regarding open data policy-making, Zuiderwijk et al. (2019) conducted an overview of the latest open data policy-making research. Their findings indicate that literature reviews concerning open data research are limited and those investigating policy-making are even rarer. Nevertheless, after analysing eight selected studies, three main topics of open data policy-making research agenda emerged. The first topic is open data policy-making and theory development, where an efficient evaluation of early practices and frameworks has been made, which indicates that the research topic of OGD is relatively new. The next topic is open data policy-making effects, which can be divided as direct and indirect and should be further examined in terms of policy formation and implementation to test their long-term behaviour. The last topic considers a multi-actor open data policy-making, where impacts on different stakeholders are examined, and comparisons of policies are presented.

Furthermore, Ingrams (2016) proposed an analytical framework for open government policy design processes by conducting a systematic literature review and

categorising open data policies. A theoretical framework rooted in open government, structuration, and policy design theory organised previous empirical findings. A conceptual model based on these theoretical guidelines was developed and tested in a case of e-participation. Prior literature already proposed conceptual models with similar factors, but none addressed open government for policy design processes in organisational theory. Eleven topics of open government research were identified in the past thirty-five years, where the most often discussed were open data, general open government, transparency, and citizen participation.

Chatfield et al. (2015) explored organisational capability challenges in transforming government through open and big data use. Four categories of big data capability challenges were identified: analytical, technical, strategical (strategic change) and socio-political, in descending order of importance. The most critical challenge is analytical capability, which describes a lack of knowledge to deal with open or big data and a shortage of analytical skills. Relatively high costs as another challenge should also not be ignored since the main aim of using big data is to minimise them. To do that, costly computational power must be provided first. Based on their classification, the authors proposed a conceptual framework for big data capability challenges, which suggests that innovative organisational culture can positively impact the alleviation of proposed challenge categories.

Yuan (2019) investigated public service co-production through citizen engagement. The author's literature review focuses on the role of information and communication technologies (ICT) in the process of co-production and its potential outcomes. Various ICT tools are used to engage citizens for co-production. Current studies show three models of ICT co-production; citizen-sourcing, automatic co-production, and government as an open platform.

Cluster B

Continuing with the second cluster and a summary of the decadal literature review of OGD by Saxena (2018), the author's findings indicate that existing literature mainly focused on theoretical and conceptual research of OGD, applied/contextual or user-focused research. Theoretical and conceptual research was primarily done on effects, barriers, comparison of initiatives, policies, and conceptual model development of OGD. In contrast, applied research focused on specific case studies in different countries and regions. Some effort was also dedicated to benchmarking different OGD initiatives and focusing on OGD portals in human interaction.

Further, Safarov et al. (2017) made a comprehensive literature review on four factors of OGD utilisation. Types, effects, conditions, and users of OGD were discussed. The authors identified eight different types of OGD utilisation, six notable effects of OGD, several conditions for implementing OGD such as quality and availability of data. Finally, users of OGD were systematically classified into different groups. The paper revealed some research gaps on OGD utilisation, principally considering the lack of empirical testing of its effects.

A research-in-progress on a case of effective implementation of OGD in the Kingdom of Bahrain (Katbi et al., 2019) explored benefits, impediments, and different assessment tools that could help governments successfully implement OGD. Their findings state that various assessment tools exist, and they vary significantly in terms of focus, scope, and area of investigation. Based on Katbi et al. (2019), it is essential to use different assessment tools as much as it is to develop new ones, where the focus should be dedicated not only to the supply but also to the demand side of OGD. Research revealed that most OGD assessment frameworks carried out had a broader

focus at the country level. Specific research focuses on a micro investigation at the level of different governmental agencies still lacks and should be further examined.

On the other hand, Huang et al. (2017) conducted a critical literature review on identifying and understanding barriers to OGD release in China. While suggesting that almost all researchers have a favourable view of OGD, their findings also indicate that barriers exist and could be classified into three main themes; institutional barriers, data integrity & quality barriers, and user participation barriers. Authors identified institutional barriers as most influential in China, where conflicts between the closed traditional and bureaucratic system of China and the OGD requirement emerged, which cannot be successfully implemented with the current system and culture remaining unchanged.

Crusoe et al. (2018) also conducted a literature review on OGD barriers from 34 articles with the purpose of better understanding the phenomenon. Their findings indicate that most barriers are rooted in technical, legal, or organisational issues, followed by participation and data-related issues. The authors proposed systematising OGD barriers in an organisational context with five distinct processes. The first three processes are identifying whether data is suitable for publishing, deciding to release the data, and publishing the data. Once the data is published, it has to be used and evaluated in terms of its impact as a fourth and fifth process.

Lastly, with a more specific scope of research, OGD initiatives, Kalampokis et al. (2011) proposed a classification scheme of OGD initiatives with two dimensions. The first dimension considers the technological aspects of OGD initiatives, and the second is domain-specific or organisational-oriented. The technological aspect is associated mainly with publishing, downloading, and processing linked data, whereas the organisational aspect considers two different approaches to data publication; direct or centralised data provision and indirect or decentralised data provision. The authors then classified twenty-four different OGD initiatives and developed a technical architecture that relies on indirect data provision.

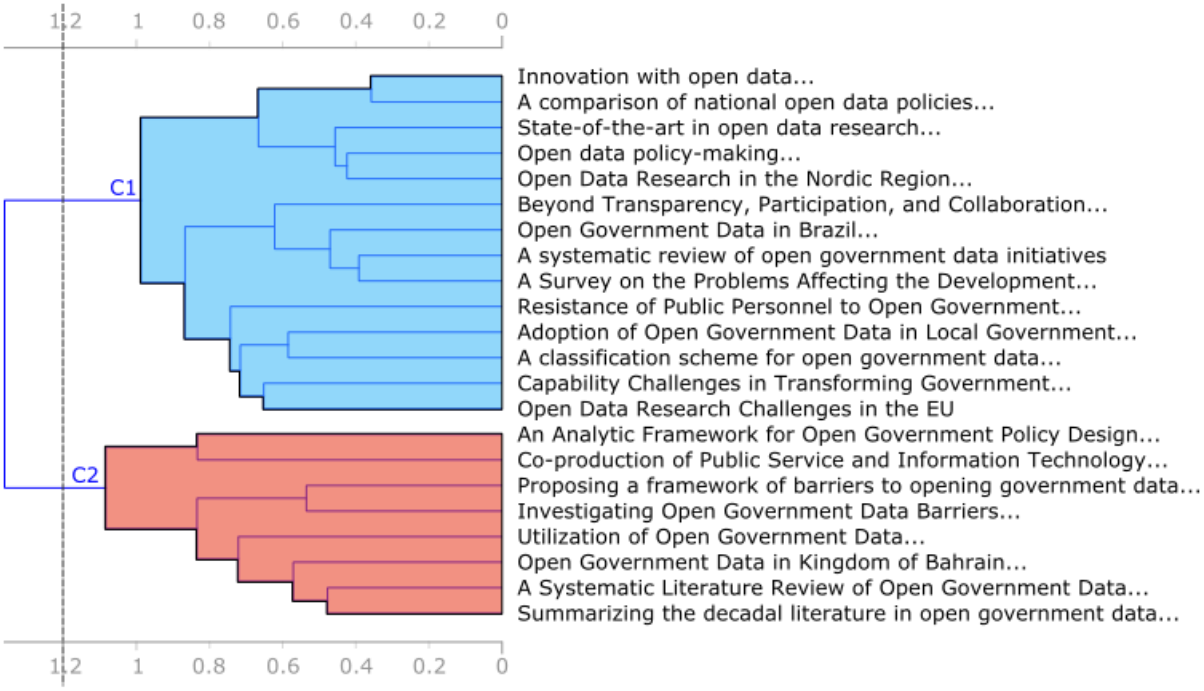
Cluster validation

To validate clusters and topics from the previous section, we performed a second round of analysis with the ScienceDirect corpus.

We used a similar preprocessing pipeline as before. A crucial addition was lemmatising the words with a UDPipe lemmatiser to compare clusters easily. Lemmatization mildly affected the original clustering, with only two papers changing cluster membership. Initial clusters were visualised with a t-SNE projection, representing a document map, where similar documents lie closer together than those that are different. We added the documents from the second corpus into the same space to compare them to the initial clusters (Figure 2).

t-SNE shows well-defined cluster regions, with the papers from the second corpus mapping to both clusters proportionally. If the new documents were entirely dissimilar to the original ones, they would be placed away from the two clusters and form a separate group.

Figure 2
Re-implementation of hierarchical clustering with added lemmatisation.

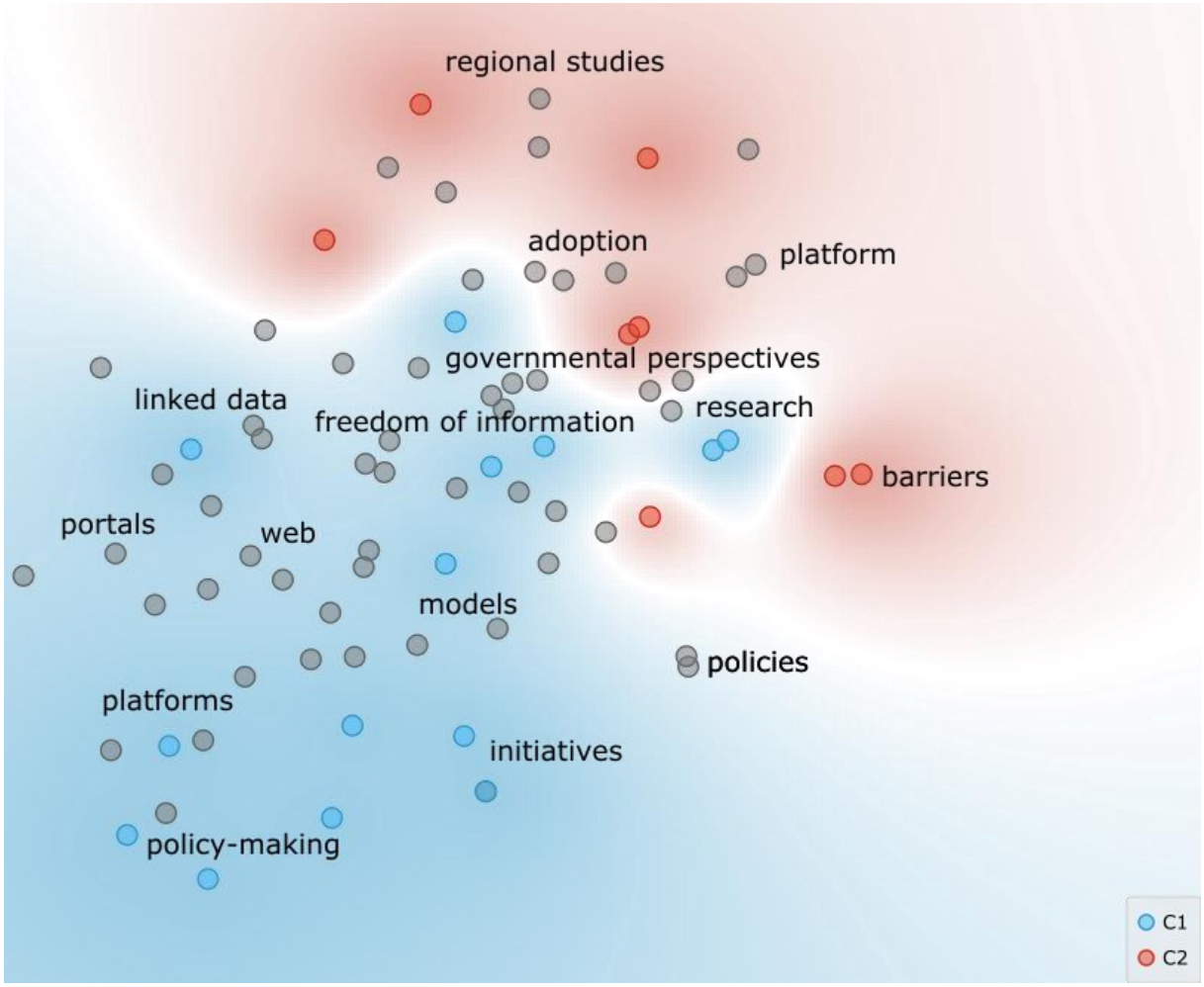


Source: Author's illustration

We defined topic regions based on the YAKE keywords (Figure 3). The blue region (cluster A) contains papers addressing government perspectives and policies on OGD, initiatives, and portals. Portals and platforms are positioned closer to the concepts of linked data, web, and models, which refer to the practical implementation of OGD. The red region (cluster B) contains papers describing regional studies, adoption of ODG, platforms, and barriers to implementation. Barriers to OGD are a slight outlier positioned closest to the research papers about OGD. This could be interpreted as barriers to ODG still being studied only at the theoretical level. To compare YAKE's topic regions with manual literature review, we can say that research topics mostly overlap with manually identified research topics, except for regional studies, which are mainly categorised into cluster A instead of cluster B. In the case of cluster A, the following research topics were identified: "Holistic approach on OGD research", "Regional context of OGD research", "OGD barriers", "OGD policy-making", "OGD and big data", and "Service co-production". The same applies to the second cluster (cluster B), where the following research topics were identified: "Research types on OGD", "OGD Utilization", "Assessment tools", and to some extent "OGD barriers".

Keyword mapping and comparison of the two corpora revealed a more intricate landscape of topics in OGD literature. In Figure 1, two papers (Yuan, 2019, Kalampokis et al., 2011) changed their clusters concerning the original clustering. Yuan's paper on the co-production of public services is an outlier, demonstrated by its position at the edge of the t-SNE landscape. Kalampokis et al. (2011) paper, on the other hand, is deeply embedded in the documents from the new corpus. Upon closer observation, the neighbourhood of documents describes linked data, portals, frameworks, and websites. This topic was not evident from the first round of clustering, proving the usefulness of an additional calibration level. The topic focuses on the practical implementation of OGD, data infrastructures, and public access to the data, all of which are essential aspects of OGD research.

Figure 3
Topic regions



Notes: Blue (cluster A) and red points (cluster B) represent papers from the original clustering. Gray points are papers from the second corpus. Points with higher similarity lie closer together. Annotations were added with the YAKE! Keyword extraction algorithm.
Source: Author's illustration

There is much overlap between approaches to OGD research in mapping the initial clusters to the landscape. Regional studies (Csáki et al., 2018; Susha et al., 2016, De Oliviera et al., 2018) are deeply embedded in a general (holistic) overview of OGD research and conducted research types, showing strong entanglement of use cases with theoretical frameworks. Barriers to OGD implementation remain one of the few distinct clusters. Keyword analysis also shows how vital societal and governmental perceptions of OGD are in implementation. In other words, barriers are intimately linked to the interplay between the benefits and risks of OGD implementation. Policy-making also forms a separate cluster at the edge of the plot, making this a distinct yet not highly represented topic. Policy-making is also closely connected to OGD frameworks; manual analysis reveals that an evaluation of early practices enables and stimulates policy-making based on open government data. Finally, the spiral intertwining blue and red cluster signify closely related topics. These documents refer to the public's vital role in accepting OGD initiatives and how public agencies implementing OGD frameworks must ensure accountability and transparency, practical and quality datasets, and uphold social values if these frameworks succeed.

Discussion and conclusions

The purpose of this study was to conduct a literature review on OGD implementation with an emphasis on previous literature reviews. As perceived in the literature, the topic is emerging and important, especially when debating transparent governance and public service innovation. We based our research on a decadal research activity between 2011 and 2020 that varies in geographical outreach and approach to research initiatives.

In the previous two sections of this research, we classified 22 articles into three different clusters with hierarchical clustering. Articles were manually reviewed to justify hierarchical clustering process categorisation and further validated with 54 articles retrieved from the ScienceDirect service. In this section, we elaborate on the findings from the literature review, interpret our results and provide a response to our intention for making this research.

The main motivation for carrying out this research was to classify existing research on open government data initiatives in terms of literature reviews to provide new and recent findings on the topic. The main research questions were: (i) RQ1: What is the main research scope of the authors?; (ii) RQ2: What are the commonalities of the various literature reviews on the subject?; (iii) RQ3: What are the latest research gaps?; and (iv) RQ4: To which groups can research on OGD implementation be classified?

As far as our research questions are concerned, we can state that a lot of research has been made;. However, not many empirical studies or design science approaches were used, most research focused on proposing different frameworks and some less on developing the platforms.

To explain the commonalities of different literature reviews on this topic and how to classify them, we used hierarchical clustering, which helped us to manually analyse articles and later either confirm the clusters created by the Orange software or make necessary corrections. As shown in Figure 1, two distinct clusters emerged, where the research focus was either on a general and general regional approach to OGD research, as presented in cluster A or on different research types on OGD, its utilisation and assessment tools, as presented in cluster B.

This study suggests that open government data received a lot of research attention and has so far been focused on many different segments of the phenomena such as impact, barriers, comparison, explanation, and benchmarking of OGD initiatives, policy-making, citizen/user interaction or participation, and the development of conceptual models for OGD development. Nevertheless, further research is needed to gain a more holistic understanding of the phenomena. These findings are also the main practical and theoretical implications of this research. Some of the proposed research directions are as follows: (i) Quantitative research to develop theories on OGD and design science research as a methodology; (ii) Holistic approach in research areas such as policy making, organisational aspects, technology and data, reuse, end users and theoretical foundations; (iii) Research on how to improve the data published by governments, open data behavioural models, and economic success factors; (iv) Research investigating legal and ethical dilemmas around open data, since the existing legal frameworks don't encourage public data use; (v) Research on the benchmarking of OGD initiatives to identify best practices for replication by others; and (vi) Research on the indirect provision of linked data in the public sector.

This study pointed out the latest research gaps and provided a comprehensive overview and classification of the topic using a method thatwasn't previously used On top of that, we validated initial clusters and put them in a wider research context by

adding another data set. While the research provided the intended results, some limitations still have to be considered. The first of these is our sample size, consisting of 22 existing literature review papers and 54 research papers. The second limitation was Orange software, which provided the intended results, but was limited by built-in similarity/dissimilarity measurements such as cosine distance measurements. As recommended in the literature (Huang, 2008), future distances for clustering text could be tested to validate our clusters, such as averaged Kullback-Leibler divergence.

References

1. Al-Anzi, F. S., AbuZeina, D. (2017), "Toward an enhanced Arabic text classification using cosine similarity and Latent Semantic Indexing", Journal of King Saud University - Computer and Information Sciences, Vol. 29 No. 2, pp. 189-195.
2. Ali Hassan, M. I., Twinomurinzi, H. (2018), "A Systematic Literature Review of Open Government Data Research: Challenges, Opportunities and Gaps", 2018 Open Innovations Conference (OI), pp. 299–304 .
3. Attard, J., Orlandi, F., Scerri, S., Auer, S. (2015), "A systematic review of open government data initiatives", Government Information Quarterly, Vol. 32 No. 4, pp. 399–418.
4. Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., Choi, H. (2014), "Big data, open government and e-government: Issues, policies and recommendations", Information Polity, Vol. 19 No. 1, pp. 5-16.
5. Boyne, G. A. (2002), "Public and Private Management: What's the Difference?" Journal of Management Studies, Vol. 39 No. 1, pp. 97-122.
6. Buchanan, C. P. (1972), "Organizational Development as a Process Strategy for Change", Educational Technology, Vol. 12 No. 10, pp. 10-14.
7. Campos, R., Mangaravite, V., Pasquali, A., Jatowt, A., Jorge, A., Nunes, C., Jatowt, A. (2020), "YAKE! Keyword Extraction from Single Documents using Multiple Local Features", Information Sciences Journal, Vol. 509, pp. 257-289.
8. Chatfield, A., Reddick, C. Al-Zubaidi, W. (2015), "Capability Challenges in Transforming Government through Open and Big Data: Tales of Two Cities", Published in ICIS 2015, available at https://pdfs.semanticscholar.org/4558/e6a25bbb045dbe0f8f9c23bc75eb20e49f4e.pdf (Mar 15 2022)
9. Crossan, M. M., Apaydin, M. (2009), "A Multi-Dimensional Framework of Organizational Innovation: A Systematic Review of the Literature", Journal of Management Studies, Vol. 47 No. 6, pp. 1154-1191.
10. Crusoe, J. Melin, U. (2018), "Investigating Open Government Data Barriers: A Literature Review and Conceptualization", Published in International Conference on Electronic Government (EGOV 2018), pp. 169-183.
11. Csáki, C., Kő, A. (2018), "Open Data Research Challenges in the EU", Lecture Notes in Computer Science, pp. 192-204.
12. Damanpour, F., Schneider, M. (2008), "Characteristics of Innovation and Innovation Adoption in Public Organizations: Assessing the Role of Managers", Journal of Public Administration Research and Theory, Vol. 19 No. 3, pp. 495-522.
13. De Oliveira, E. F., Silveira, M. S. (2018), "Open government data in Brazil a systematic review of its uses and issues", Proceedings of the 19th Annual International Conference on Digital Government Research Governance in the Data Age - Dgo '18, pp. 1-9.
14. Demšar, J., Zupan, B., Leban, G., Curk, T. (2004), "Orange: From Experimental Machine Learning to Interactive Data Mining", Knowledge Discovery in Databases: PKDD 2004, pp. 537-539.
15. Elliott, I. C. (2020), "Organisational learning and change in a public sector context", Teaching Public Administration, Vol. 38 No. 3, pp. 270-283.
16. 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, pp. 1-20.

17. Haini, S. I., Rahim, N. Z. A., Zainuddin, N. M. M. (2019), "Adoption of Open Government Data in Local Government Context: Conceptual Model Development", Proceedings of the 2019 5th International Conference on Computer and Technology Applications - ICCTA 2019, pp. 193-198.
18. Harrison, T. M., Guerrero, S., Burke, G. B., Cook, M., Cresswell, A., Helbig, N., … Pardo, T. (2011), "Open Government and e-government: Democratic Challenges From a Public Value Perspective", in Proceedings of the 12th Annual International Digital Government Research Conference on Digital Government Innovation in Challenging Times - Dg.o '11, pp. 83-97.
19. Hossain, M. A., Dwivedi, Y. K., Rana, N. P. (2015), "State-of-the-art in open data research: Insights from existing literature and a research agenda", Journal of Organizational Computing and Electronic Commerce, Vol. 26 No. 1-2, pp. 14-40.
20. Huang, A. (2008), "Similarity Measures for Text Document Clustering", Proceedings of the New Zealand Computer Science Research Student Conference (NZCSRSC) 2008, pp. 49-56.
21. Huang, R., Lai, T., Zhou, L. (2017), "Proposing a Framework of Barriers to Opening Government Data in China: A Critical Literature Review", Library Hi Tech, Vol. 35 No. 3, pp. 421-438.
22. Ingrams, A. (2016), "An Analytic Framework for Open Government Policy Design Processes", in Proceedings of 15th IFIP WG 8.5 International Conference, pp. 203-214.
23. Jetzek, T., Avital, M. Bjørn-Andersen, N. (2013), "Generating value from open government data", in Proceedings of International Conference on Information Systems (ICIS 2013): Reshaping Society Through Information Systems Design, available at https://dblp.org/db/conf/icis/icis2013.html (Feb 1 2022)
24. 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, Vol. 6 No. 3, pp. 266-285.
25. Kassen, M. (2013), "Globalisation of e-government: open government as a global agenda; benefits, limitations and ways forward", Information Development, Vol. 30 No. 1, pp. 51-58.
26. Katbi, A.-K., Al-Ammary, J. (2019), "Open Government Data in Kingdom of Bahrain: Towards an Effective Implementation Framework", New Knowledge in Information Systems and Technologies, pp. 699-715.
27. Kitchenham, B. (2004), "Procedures for performing systematic reviews", Keele, UK, Keele University, Vol. 33, pp. 1-26.
28. Lewin, K. (1947), "Group decision and social change", in Newcomb T. M., Hartley E. L. (Eds.), Readings in Social Psychology, New York, Henry Holt, pp. 330-344.
29. Lin, Y.-S., Jiang, J.-Y., Lee, S.-J. (2014), "A Similarity Measure for Text Classification and Clustering", IEEE Transactions on Knowledge and Data Engineering, Vol. 26 No. 7, pp. 1575-1590.
30. Lourenço, R. P. (2015), "An analysis of open government portals: A perspective of transparency for accountability", Government Information Quarterly, Vol. 32 No. 3, pp. 323-332.
31. Marčeta, M. Bojnec, Š. (2020), "Drivers of Global Competitiveness in the European Union Countries in 2014 and 2017", Organizacija Journal, Vol. 53 No. 1, pp. 37-52.
32. Maxwell, A., Osae-Larbi, A. J. (2015), "Organization Development Models: A Critical Review and Implications for Creating Learning Organizations", European Journal of Training and Development Studies, Vol. 2 No. 3, pp. 29-43.
33. Moon, M. J. (2002), "The Evolution of E-Government among Municipalities: Rhetoric or Reality?", Public Administration Review, Vol. 62 No. 4, pp. 424-433.
34. Murtagh, F., Contreras, P. (2011), "Algorithms for hierarchical clustering: an overview", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 2 No. 1, pp. 86-97.
35. Nam, T. (2012), "Citizens' attitudes toward Open Government and Government 2.0", International Review of Administrative Sciences, Vol. 78 No. 2, pp. 346-368.
36. Nugroho, R., Zuiderwijk, A., Janssen, M. de Jong, M. (2015), "A comparison of national open data policies: lessons learned", Transforming Government: People, Process and Policy, Vol. 9 No. 3, pp. 286-308.

37. Peters, T., Watterman, H. R. (1982), In Search of Excellence, HarperCollins Publishers.
38. Pettigrew, A. M. (1973), The Politics of Organisational Decision Making, London, Tavistock.
39. Roa, H. N., Loza-Aguirre, E., Flores, P. (2019), "A Survey on the Problems Affecting the Development of Open Government Data Initiatives", in 2019 Sixth International Conference on eDemocracy eGovernment (ICEDEG), pp. 157-163.
40. Safarov, I., Meijer, A., Grimmelikhuijsen, S. (2017), "Utilisation of open government data: A systematic literature review of types, conditions, effects and users", Information Polity, Vol. 22 No. 1, pp. 1-24.
41. Saif, H., Fernández, M., He, Y. Alani, H. (2014), "On stopwords, filtering and data sparsity for sentiment analysis of Twitter", in LREC 2014, Ninth International Conference on Language Resources and Evaluation, Proceedings, pp. 810-817.
42. Saxena, S. (2018), "Summarising the decadal literature in open government data (OGD) research: a systematic review", Foresight, Vol. 20 No. 6, pp. 648-664.
43. Sprecher, M. H. (2000), "Racing to E-government: Using the Internet for Citizen Service Delivery", Government Finance Review, Vol. 16, pp. 21-22.
44. Susha, I., Johannesson, P., Juell-Skielse, G. (2016), "Open Data Research in the Nordic Region: Towards a Scandinavian Approach?", Electronic Government, pp. 61-73.
45. Tolbert, C. J., Mossberger, K. (2006), "The effects of E-government on trust and confidence in government", Public administration review, Vol. 66 No. 3, pp. 354-369.
46. Tsoukas, H. Chia, R. (2002), "On Organizational Becoming: Rethinking Organizational Change", Organization Science, Vol. 13 No. 5, pp. 567-582.
47. Van de Ven, A. H., Poole, M. S. (1995), "Explaining Development and Change in Organisations", The Academy of Management Review, Vol. 20 No. 3, pp. 510-540.
48. Van der Maaten, L. Hinton G. (2008), "Visualizing Data using t-SNE", Journal of Machine Learning Research, Vol. 9 No. 86, pp. 2579-2605.
49. Vetrò, A., Canova, L., Torchiano, Orozco Minotas, C., Iemma, R., Morando, F. (2016), "Open data quality measurement framework: Definition and application to Open Government Data", Government Information Quarterly, Vol. 33 No. 2, pp. 325-337.
50. Walker, R. M., Andrews, R. (2013), "Local Government Management and Performance: A Review of Evidence", Journal of Public Administration Research and Theory, Vol. 25 No. 1, pp. 101-133.
51. Webster, J., Watson, R. T. (2002), "Analysing the past to prepare for the future: writing a literature review", MIS Quarterly, Vol. 26 No. 2, pp. 13-23.
52. Weick, K. E., Quinn, R. E. (1999), "Organizational Change and Development", Annual Review of Psychology, Vol. 50 No. 1, pp. 361-386.
53. Welle Donker, F., van Loenen, B. (2017), "How to assess the success of the open data ecosystem?", International Journal of Digital Earth, Vol. 10 No. 3, pp. 284-306.
54. West, D. M. (2004), "E-Government and the Transformation of Service Delivery and Citizen Attitudes", Public Administration Review, Vol. 64 No. 1, pp. 15-27.
55. Wirtz, B. W., Piehler, R., Thomas, M.-J., Daiser, P. (2015), "Resistance of Public Personnel to Open Government: A cognitive theory view of implementation barriers towards open government data", Public Management Review, Vol. 18 No. 9, pp. 1335-1364.
56. Yuan, Q. (2019), "Co-production of Public Service and Information Technology", in Proceedings of the 20th Annual International Conference on Digital Government Research (dg.o 2019), pp. 123-132.
57. Zuiderwijk, A., Hinnant, C. C. (2019), "Open data policy-making: A review of the state-of-the-art and an emerging research agenda", Information Polity, Vol. 24, No. 2, pp. 1-13.
58. Zuiderwijk, A., Janssen, M., Davis, C. (2014), "Innovation with open data: Essential elements of open data ecosystems", Information Polity, Vol. 19 No. 1, pp. 17-33.

About the authors

Aljaž Ferencek is a doctoral student at the Faculty of Organizational Sciences at the University of Maribor. He received his master's degree in the same faculty from information systems studies. His research interests include data science and open and open government data, and he has already published research on this subject. The author can be contacted at **aljaz.ferencek1@student.um.si**

Mirjana Kljajić Borštnar is a full professor of information systems at the University of Maribor, Faculty of Organizational Sciences. Her research work covers decision support, data mining and organisational learning. She is active in program committees of Bled eConference, Symposium on OR in Slovenia, Dnevi slovenske informatike and DataScience. She is an academic representative in Strategic Research and Innovation Partnership - Smart cities and communities for AI, HPC & Big Data; executive board member of Ai4Slovenia; and editor in chief of the journal Uporabna informatika. The author can be contacted at **mirjana.kljajic@um.si**

Ajda Pretnar Žagar holds a PhD in anthropology from the University of Ljubljana. She is a researcher at the Laboratory for Bioinformatics, the Faculty of Computer and Information Science, and the Institute of Contemporary History. Her research focuses on the methodology of interdisciplinary and multidisciplinary research and the uses of machine learning and data mining in the humanities and social sciences. She teaches at the Higher School of Economics in Moscow, Russia. The author can be contacted at **ajda.pretnar@fri.uni-lj.si**