Enhancing transparency through open government data: the case of data portals and their features and capabilities

Martin Lněnička, Renata Machova, Jolana Volejníková, Veronika Linhartová, Radka Knezackova, and Miloslav Hub

Abstract

Purpose – The purpose of this paper was to draw on evidence from computer-mediated transparency and examine the argument that open government data and national data infrastructures represented by open data portals can help in enhancing transparency by providing various relevant features and capabilities for stakeholders' interactions.

Design/methodology/approach – The developed methodology consisted of a two-step strategy to investigate research questions. First, a web content analysis was conducted to identify the most common features and capabilities provided by existing national open data portals. The second step involved performing the Delphi process by surveying domain experts to measure the diversity of their opinions on this topic.

Findings – Identified features and capabilities were classified into categories and ranked according to their importance. By formalizing these feature-related transparency mechanisms through which stakeholders work with data sets we provided recommendations on how to incorporate them into designing and developing open data portals.

Social implications – The creation of appropriate open data portals aims to fulfil the principles of open government and enables stakeholders to effectively engage in the policy and decision-making processes.

Originality/value – By analyzing existing national open data portals and validating the feature-related transparency mechanisms, this paper fills this gap in existing literature on designing and developing open data portals for transparency efforts.

Keywords Transparency, Content analysis, Delphi method, Open government data, Open data portals, Features and capabilities

Paper type Research paper

1 Introduction

The concept of open government can be viewed as a multilateral, political, and social process, in which actions by government take place (Wirtz and Birkmeyer, 2015), or as a set of dimensions (Gil-Garcia *et al.*, 2020). Both these views require participation and collaboration of citizens by using modern Information and Communication Technologies (ICT) to improve information availability and transparency of governmental actions. Government transparency is interpreted as a way of creating openness through disclosing information about governmental actions and processes (Lindstedt and Naurin, 2010). This principle helps in reducing corruption, improving government effectiveness, and efficiently allocating resources (Bertot *et al.*, 2010; Bogdanović-Dinić *et al.*, 2014; Gil-Garcia *et al.*, 2020; Janssen *et al.*, 2017; Moon, 2020; Murillo, 2015). The widespread availability and use of ICT through computerized systems in the public sector agencies and institutions is still a key tool (channel) to enhance transparency (Meijer, 2009). It is further boosted by the rapid growth of Internet access to most people, making the supply of online information more competitive and accessible (Goel *et al.*, 2012). Nowadays, technological developments in the world heavily rely on data and the usage of ICT

in the public sector can benefit in various ways (Lourenço *et al.*, 2017).

Due to these shifts, the opening of data became a global phenomenon (Bertot *et al.*, 2010; Bogdanović-Dinić *et al.*, 2014; Glassey, 2017; Veljković *et al.*, 2014). From 2010 onward, open data have been linked into broader open government reforms (Noveck, 2017), which have in turn tied open data into public management reforms (Clarke and Margetts, 2014). According to Moon (2020), the main shift in open government initiatives is represented by the change of citizens' role from passive informed and service recipients to active co-producers of public services and users of open data. Data transparency refers to opening public sector information and enabling citizens and other stakeholders to access government-held data in a uniform way (Veljković *et al.*, 2014). Open Government Data (OGD) became one of the key concepts shaping the efforts of governments to be more transparent and accountable about the use of public resources (Attard *et al.*, 2015; Kimball, 2011; Lindstedt and Naurin, 2010; Noveck, 2017; Safarov *et al.*, 2017). Besides that, making data open can provide greater returns on public investment, help policymakers address complex problems, creation of steady trust and credibility in government, improve public policies and the efficiency and quality of public services, cost reduction, etc. (Bogdanović-Dinić *et al.*, 2014; Gil-Garcia *et al.*, 2020; Matheus and Janssen, 2020).

In the past few years, various authors have reported that the release of OGD on open data portals seems to affect processes that have contribution to transparency in areas where corruption, wastage and inefficiency happen the most (Bertot *et al.*, 2010; Hogan *et al.*, 2017; Lourenço, 2015; Matheus and Janssen, 2020; Murillo, 2015; Safarov *et al.*, 2017). According to Klein *et al.* (2018), "*OGD portals need to address a number of mechanisms so that society can effectively discover, extract, and utilize the data.*" Although the perspective of transparency achieved through data portals was already analysed by Lourenço (2015), Klein *et al.* (2018) or Thorsby *et al.* (2017), these authors dealt mostly with the basic principles and characteristics of open data. There is a research gap in identifying the enabling features that are provided by open data portals and are crucial to work with and enhance transparency through OGD.

In this paper, we argue that opening, sharing, and reusing OGD through open data portals should be done more effectively to improve government transparency. The main assumption is that opening, sharing, and reusing these data through a data infrastructure represented by an open data portal will facilitate a better management of public sector information delivery to citizens and other stakeholders. Our argument is based on the case of existing open data portals and their role as a tool to provide greater transparency of government programs, activities, publications, and spending. Additionally, comparing unique features and capabilities of open data portals will contribute to the understanding of key processes and resources that may help to develop more targeted strategies and activities regarding accessing to appropriate government data. We define a feature as "*functionality or activity provided by the open data portal through which various stakeholders are able to work with data sets for their intended purpose.*"

This paper exploits two questions about the potential of open data portals to enhance transparency. Our research is focused on the following research questions (RQ):

RQ1: What are the most common features and capabilities of open data portals and how they group into relevant transparency-enhancing mechanisms?

RQ2: What is their relative importance assigned by domain experts on enhancing transparency and how can be these findings used to develop open data portals?

We adopt a research methodology consisting of a two-step strategy to investigate these questions. First, a web content analysis is conducted to identify the most common features and capabilities provided by existing national open data portals. The second step involves performing the Delphi process by surveying domain experts to measure the diversity of their opinions on this topic (Okoli and Pawlowski, 2004). It allows us to determine the relative importance of the preferred features and capabilities by the experts. This research contributes scientifically by being the first to provide insights into feature-related transparency mechanisms provided by open data portals and their ability to stimulate the reuse of OGD through unique features and capabilities. In response to a study by Klein *et al.* (2018), who explored the transparency mechanisms as "*procedures, artifacts, or a set of actions that aim for transparency by respecting principles of OGD*", we conceptualize them on the basis of existing open data portals that are critical component of data infrastructures and represent the entry point for reusing OGD. Since our aim is to explore open data portals as a complex whole we do not strictly focus only on the characteristics that OGD should meet but also on other feature-related transparency mechanisms through which it is enabled to work with OGD.

Based on our results, we have formulated recommendations on how public sector agencies and institutions should be put into practice on developing open data portals as a tool to increase transparency. The launch of open data portals as a central online hub for public sector information represents a platform with features and capabilities enabling engagement of stakeholders. Moreover, information gathered through the comparison allows us to make connections between stakeholders with their different knowledge and skills and relevant features and capabilities needed to gain greater transparency impact. In order to achieve these, the design of open data portals has to follow a strategy based on improving availability of appropriate features and capabilities around data sets instead of opening more and more data.

Following this introduction, this paper is organized into an additional five sections. Section two explains the research methodology used for conducting our study. Section three presents benchmarking framework for contextualizing feature-related transparency mechanisms provided by open data portals. It is followed by results and recommendations. Section five discusses implications and limitations of the study. Finally, we draw conclusions.

2 Research methodology

This study is conceptually grounded in computer-mediated transparency characterized by Meijer (2009). It deals with providing information to citizens in more innovative ways through computerized systems and Internet. We contribute to this concept by describing how features of open data portals should be designed to ensure that stakeholders are capable of working with the information that has been made available online. According to Lourenço (2015) and Lourenço *et al.* (2017), government transparency is nowadays associated with Internet-driven ICT, the open data movement and with open government portals. In this context, Veljković *et al.* (2014) distinguish between government transparency and data transparency. Ojo *et al.* (2016) extended this approach in the context of open data portals and argued that the portals' features should explicitly mediate effective transparency in terms of access and understandability. In this regard, we further follow the approach of Ojo *et al.* (2016) and evaluate the access as "*features available to work with OGD on the data portal*" and the understandability as "*features take into account stakeholders with different level of skills*".

The research methodology has been developed in order to answer the research questions, see Figure 1. Because the study is concerned with understanding to what extent users can effectively

work with OGD on open data portals, corresponding features and capabilities implemented for this purpose are crucial. Thus, a web content analysis (feature analysis) approach is used to scan through the selected national open data portals in order to identify them. To maintain a high reliability and validity of the approach, a literature study of relevant empirical studies concerning open data portals was conducted, including Alexopoulos *et al.* (2017), Attard *et al.* (2016), Chatfield and Reddick (2017), Klein *et al.* (2018), Lourenço (2015), Máchová and Lněnička (2017), Máchová *et al.* (2018), Nikiforova and McBride (2021), Ruijer *et al.* (2017), Thorsby *et al.* (2017) and Zuiderwijk *et al.* (2013). It is important to point out that open data portals on the national level are compared since they provide rich features and capabilities. The list of them presented in Máchová and Lněnička (2017) was used as a basis and updated with portals that were launched after 2017. From a total number of 110 portals found a subset of 82 portals available in English was selected for our study.

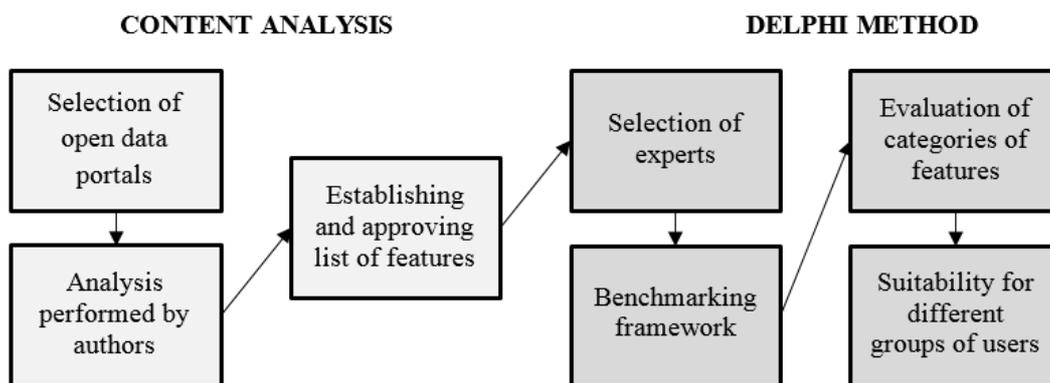**CONTENT ANALYSIS**   **DELPHI METHOD**

Figure 1. Steps of the research methodology

Authors scrutinized the interfaces of targeted open data portals to find out what features and capabilities are implemented to enable users work with OGD. They evaluated them from a data user perspective. This included reading information about the portal and other documentation, inspecting menu options, navigation links, search options, forms, and feedback instruments, etc. While most of the features such as list of data sets, data search and filter or data categories could be easily identified on each portal's homepage, the other features required thorough search through all the web pages. Based on this exploration, the authors evaluated whether each feature is suitable for the study purposes and unique in contrast to features found on other portals. Total number of features approved by authors was 48. These were listed in a table and supplemented by notes about how each feature is supported. It should be noted that we did not seek to evaluate the existence or lack of a feature on each portal, but we aim to create a comprehensive view of features and capabilities that are provided by the portals globally.

Afterwards, the Delphi method is used to classify the features into categories (feature-related transparency mechanisms) that form the conceptual model serving as a benchmarking framework for the evaluation of the relative importance. The main reason for the use of this method is the lack of other research evidence on the topic. Further arguments include: 1) it does not need to bring the experts together and enables to involve more experts with a broader interdisciplinary background, 2) it affords anonymity to participants and privacy for iteration and to change one's mind over several rounds and 3) it facilitates the transformation of opinions into consensus. The steps of this method follow the guidelines of Okoli and Pawlowski (2004). The required qualifications of experts include a minimum of three years' experience in the field of the research problem and at least M.Sc. A group of ten experts have been specially selected for their expertise and interdisciplinary background that

is relevant for OGD and transparency efforts. It enables the opportunity to check the validity of the cross-disciplinary nature of the issue (Grisham, 2009).

Experts' qualifications are summarized in Table 1, including a job position, academic title, years of experience, and professional expertise. Our aim was to involve experts that represent a range of opinions about the research topic and can evaluate needs of data users with different level of skills. All the experts are familiar with the Delphi process and agreed to participate in all Delphi rounds needed. Their motivation as well as the quality of the instructions are important since a high dropout rate among experts may lead to non-representative results (Grisham, 2009; Okoli and Pawlowski, 2004). Ensuring that the process is robust, consistent, transparent, and meets the quality aspects arising from the previous uses of the Delphi method determines the external validity of the results.

Table 1. Experts' qualifications participated in the Delphi process

Job position	Academic title	Experience	Expertise
Academic staff	Assoc. Prof.	> 20 years	Economics, management and corruption
Academic staff	Assoc. Prof.	10 - 15 years	Computer science, engineering and usability
Academic staff	Ph.D.	15 - 20 years	Information systems and e-government
Academic staff	Ph.D.	10 - 15 years	Administration and economics of corruption
Independent researcher	Ph.D.	10 - 15 years	Computer science and pedagogy
Public official	Ph.D.	10 - 15 years	Economics, administration and finance
Public official	M.Sc.	> 20 years	Administration and public economics
Public official	M.Sc.	5 - 10 years	Law, legislation, and public policy
Analyst	M.Sc.	15 - 20 years	Quantitative methods and computer science
Analyst	M.Sc.	5 - 10 years	Digital marketing, communication and media

The feature-related transparency mechanisms are evaluated by the experts from a data user perspective. The list of 48 features is used to construct the online survey instrument distributed in subsequent quantitative rounds. A five-point Likert scale is applied for this purpose. After each of these rounds, the responses are provided to all 10 panel members and they are asked to decide whether they want to change any of their responses. In addition, they are also asked to discuss the reasons for their choice. The subsequent round questionnaire is constructed from the results gathered from the previous one. The process is finished after no one wants to change their responses. In most cases, the Delphi process is limited to a maximum of five rounds. However, only three rounds were realized to reach a consensus on the categories – 14 in the first round, 11 in the second round, and the final 9 categories in the third round. The scores are analysed after each round to calculate mean values and standard deviations of each question. Mean value indicates the central tendency of feedback, whereas standard deviation shows the achievement degree of consensus. The reliability is measured using Cronbach's alpha coefficient.

After the completion of the consensus process in which the features were grouped into respective categories of mechanisms, the experts were asked to assign relative importance to each category of features as well as types of data users. The five-point Likert scale (from 1 = extremely unimportant to 5 = extremely important) is utilized to determine the suitability of selected feature-related transparency mechanisms including suitability for different groups of users regarding their skills. This perspective deals with requirements of different roles in which data users may interact with the portal (Alexopoulos *et al.*,2017; Attard *et al.*, 2016; Ruijer *et al.*, 2017), i.e., basic,

intermediate, and advanced. Finally, for the purpose of this study, the levels of expected skills are defined based on International Certification of Digital Literacy (ICDL) profiles as follows:

Beginner – it is expected that the user can do basic browsing, text entry, searching and filtering content, and downloading a file.

Advanced – it is expected that the user can fill in and submit a form, attach a file, work with visualizations, differentiate between various file formats, and pre-process data sets.

Expert – it is expected that the user can use an Application Programming Interface (API), work with interactive visualizations including external applications, process, and link data from different sources, analyse data sets, and publish and share own resources.

3 A benchmarking framework for contextualizing feature-related transparency mechanisms provided by open data portals

Difficulty in measuring impact generated by open data portals is discussed in many papers. A number of benchmarking frameworks and metrics was introduced to provide a way of understanding the most important features and capabilities (Chatfield and Reddick, 2017; Hogan *et al.*, 2017; Máchová and Lněnička, 2017; Thorsby *et al.*, 2017; Welle Donker and van Loenen, 2017). Assessment frameworks such as benchmarking and comparison between different open data portals are used to identify best practices in design and development (Welle Donker and van Loenen, 2017). According to Lourenço *et al.* (2017), these frameworks are a necessary part of the OGD driven public accountability and transparency efforts. To quantify the effects, Sayogo *et al.* (2014) applied web content analysis (content analysis technique) and proposed a framework for benchmarking OGD efforts and concluded that there are different stages of ODG portal development. The framework of Zuiderwijk *et al.* (2013) compared features dealing with creating, opening, finding, using, and discussing open data. Afful-Dadzie and Afful-Dadzie (2017) audited the OGD web portals and investigated the preferred and desirable qualities an OGD portal should possess. Thorsby *et al.* (2017) introduced these categories of features: content, help features, policy, and results. As reported by Attard *et al.* (2015), published data properties, features of portals, participation of stakeholders, maturity of initiatives, and feedback of stakeholders are the most important aspects in open government initiative evaluations.

This topic is closely related to the usability concept since the chance that data are discovered may increase if these are published in an accessible and user-friendly portal (Máchová *et al.*, 2018; Nikiforova and McBride, 2021; Welle Donker and van Loenen, 2017). Regarding the usability needs, portals should facilitate a broad range of features corresponding to various needs and requirements of different roles in which stakeholders use the portal (Ruijer *et al.*, 2017). Although different value creating roles can be identified in this ecosystem (Attard *et al.*, 2016; Lněnička and Komárková, 2019), there are two distinct roles in which stakeholders are interacting with open data portals, namely a data provider and a data user (Máchová *et al.*, 2018). However, data users are limited to a level of skills that are needed to interact with portal features and capabilities (Attard *et al.*, 2016).

By analysing these studies there is still no clear cut-off line between enabling mechanisms represented by features and characteristics of OGD represented by open data principles. Most of them do not take into consideration those two different issues and makes it difficult to clearly identify the mechanisms that may contribute to increasing transparency. In order to clarify the mechanisms that occur around open data portals, the conceptual model in Figure 2 is proposed. Promoting OGD and open data portals efforts should ensure that stakeholders will be able to engage in public control and open government. A usable platform for engagement of stakeholders should provide such features

and capabilities that enable to effectively work with data sets. It considers different dimensions (Gil-Garcia *et al.*, 2020; Lněnička and Komárková, 2019) through which stakeholders can be engaged. For both roles, different skills, knowledge, and abilities affect how data will be publish and reused. We explore more deeply the expected skills of data users. These must be considered while designing features on open data portals because some of them are more advanced and their use may require additional skills and experience. The portals should provide features for different groups of users so that no one is left out.
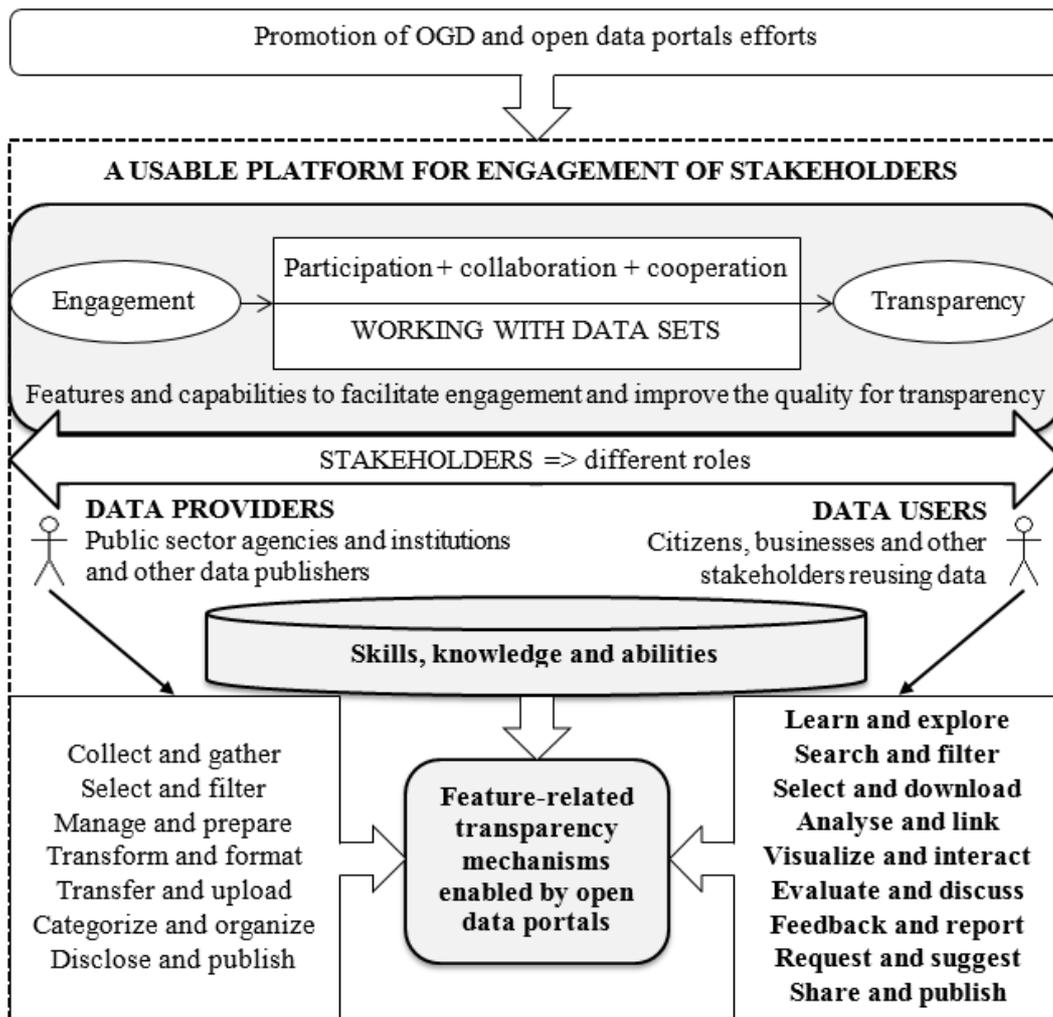

Figure 2. Clarification of feature-related transparency mechanisms that occur around open data portals

The second conceptual model further elaborates the feature-related transparency mechanisms provided by open data portals from a data user perspective. As can be seen in Figure 3, nine groups of features were established. Each of the groups contains summarized features and capabilities that were found on at least one portal. It appears to be appropriate and useful, as it concentrates on all steps of the OGD lifecycle, including the exploring, searching, downloading, analysing, etc., allowing data users to work with OGD. Increased transparency is then achieved through OGD that are reused in the various phases of the OGD lifecycle.
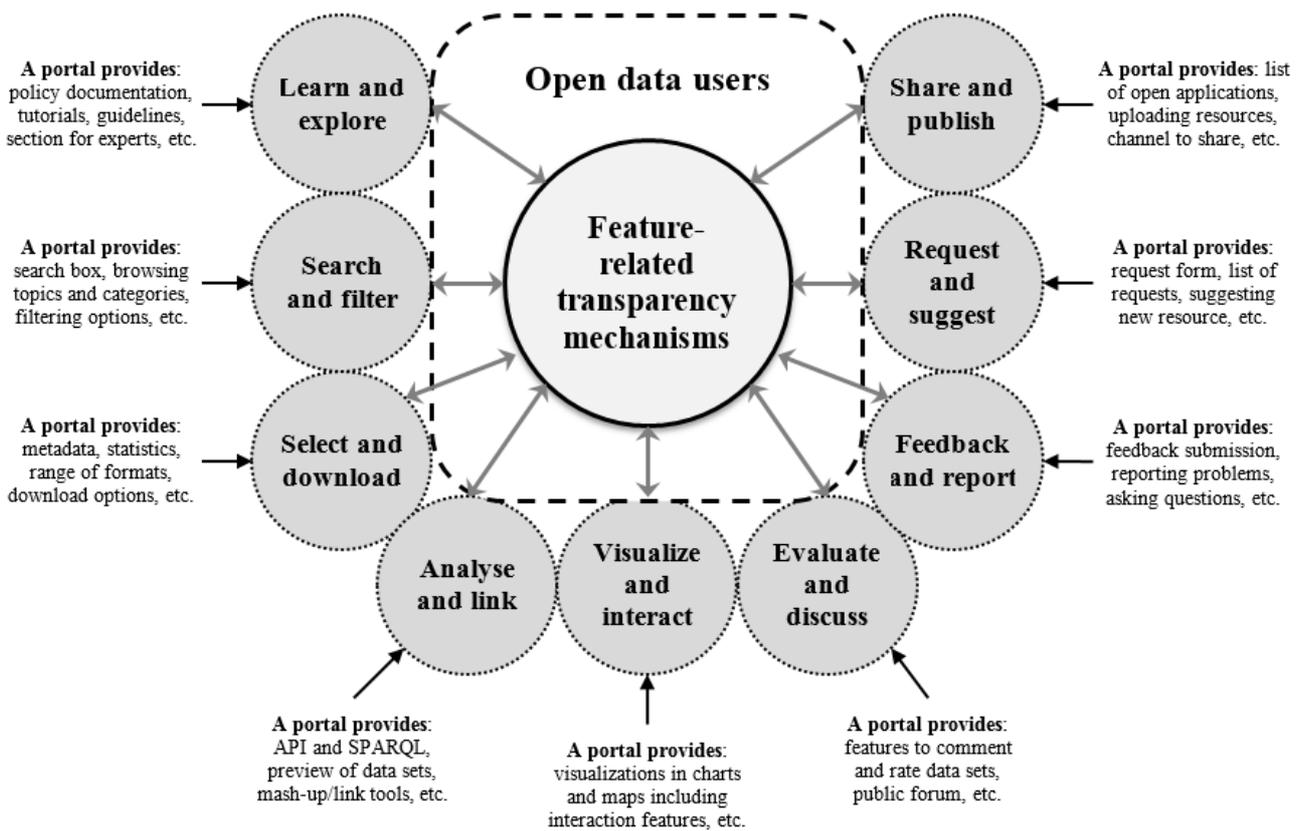
Figure 3. Classification of transparency mechanisms provided by open data portals from a data user perspective

4 Results and recommendations

The reliability of the questionnaire was measured using Cronbach's alpha coefficient to determine how closely related a set of features are as a group. A removal of an inappropriate category increases the alpha. The SPSS 23 software was used, and the values were computed independently for all three rounds of the Delphi process. A Cronbach alpha of 0.350 for all categories was obtained for round one, 0.518 for round two and 0.652 for round three. For exploratory research, values of 0.6–0.7 are acceptable (Gliem and Gliem, 2003). After the consensus process was complete, the features were grouped into respective categories of mechanisms, which can be seen on the conceptual model in the previous section, and the RQ1 was fulfilled.

Regarding the RQ2, the experts were asked to assign relative importance to each mechanism. The results in Figure 4 show that the most important categories of features are those related to search and filter together with select and download. Their relevance lies in the fact that even stakeholders with basic computer skills can use them. These are followed by features enabling to learn and explore the capabilities of the portal and the ways how OGD can be reused. At the same level, features to visualize and interact with open data sets and provide feedback or report a problem about them are listed. Although other features were also evaluated as important for the transparency efforts, their using requires at least advanced skills and knowledge to reuse OGD.
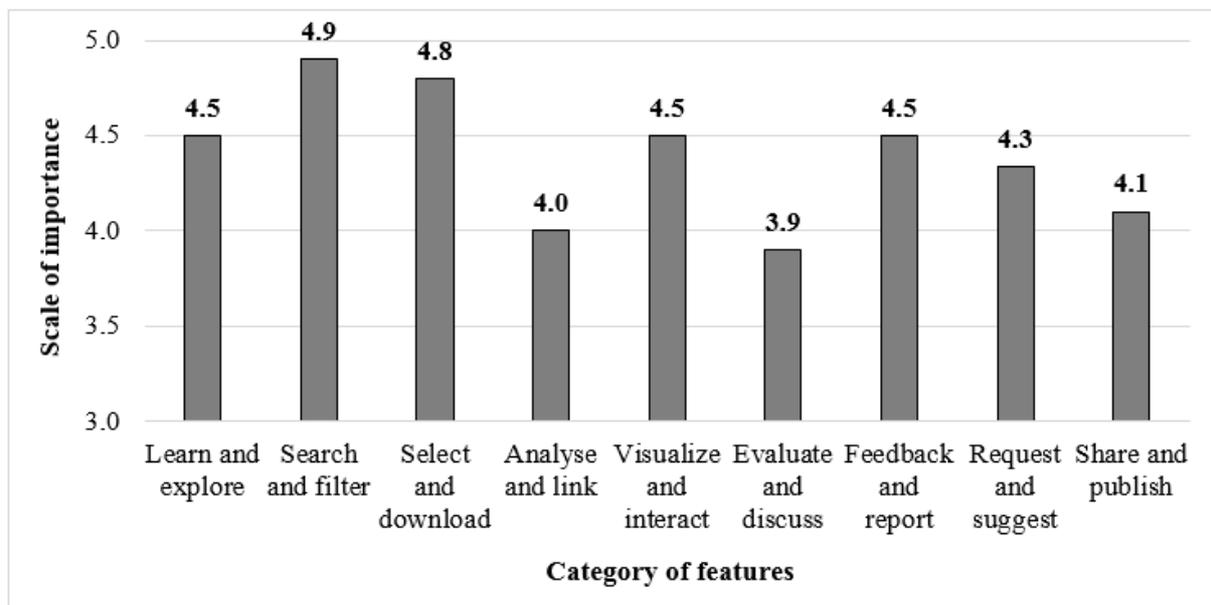
Figure 4. The importance of each category for transparency efforts

Then, each expert was asked to rank these categories. Figure 5 shows the mean rank for each category of features, where lower number means the rank is more important. The figure better illustrates the differences between categories.
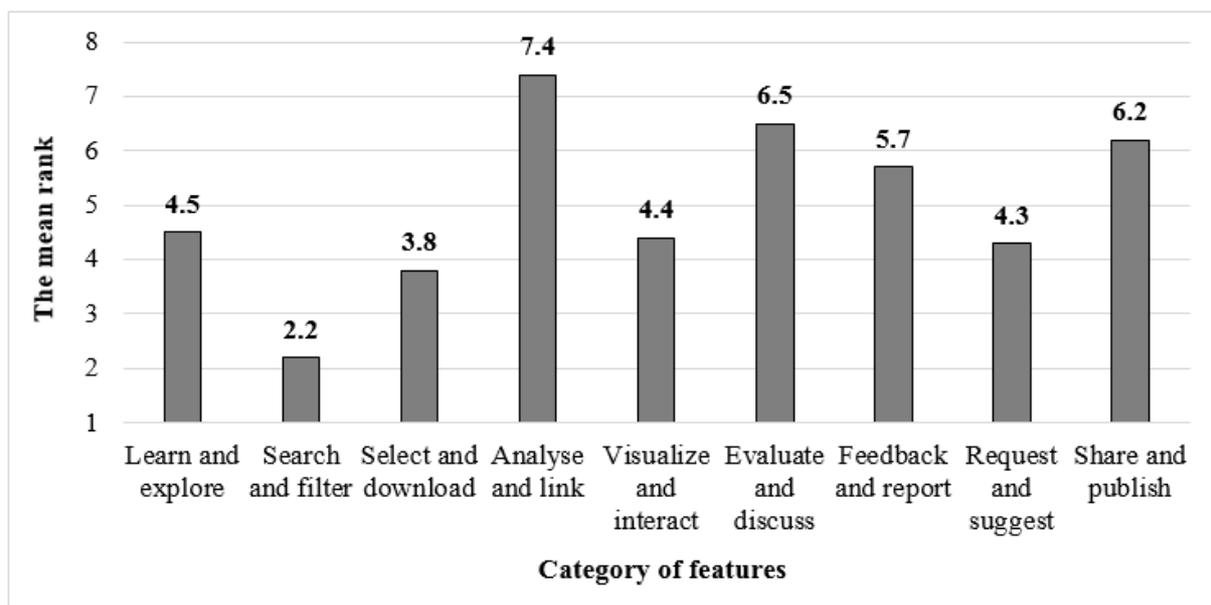

Figure 5. The mean rank for each category of features

Because of the need to deal with different level of skills and data literacy among data users, we asked the experts to evaluate the suitability of identified categories of features for different groups of users (beginner, advanced, and expert). They got instructions and guidelines about the general level of required skills based on the concept of ICDL, which can be applied to work with online information and open data portals. They evaluated the suitability from the point of view of their expertise, enabling us to discuss results through interdisciplinary perspective.

As can be seen in Figure 6, most of the features provided by open data portals should consider these categories as relevant for transparency efforts to be successful. A difference between groups of

users with advanced and expert skills is less significant that in the case of users with basic skills. The educated person is more likely to be interested in the issue, will be able to find and work with these data and then ask questions, search for answers, make the appropriate conclusions, and to publish and disseminate them. In this regard, governments should put emphasis on education and stakeholders' interest in the issue, invest in discoverability best practices, and promote the use of OGD through training hackathons, workshops, consultations, and collaborations with researchers or community groups. Beginners should be more interested in public affairs, which is the question of their interest in education that will be higher, the more it will be necessary for everyday life to use open data. Thus, the issue of economic literacy should already be addressed to the level of primary and secondary education (similarly as financial literacy of the population). Increased transparency is associated with these prerequisites.
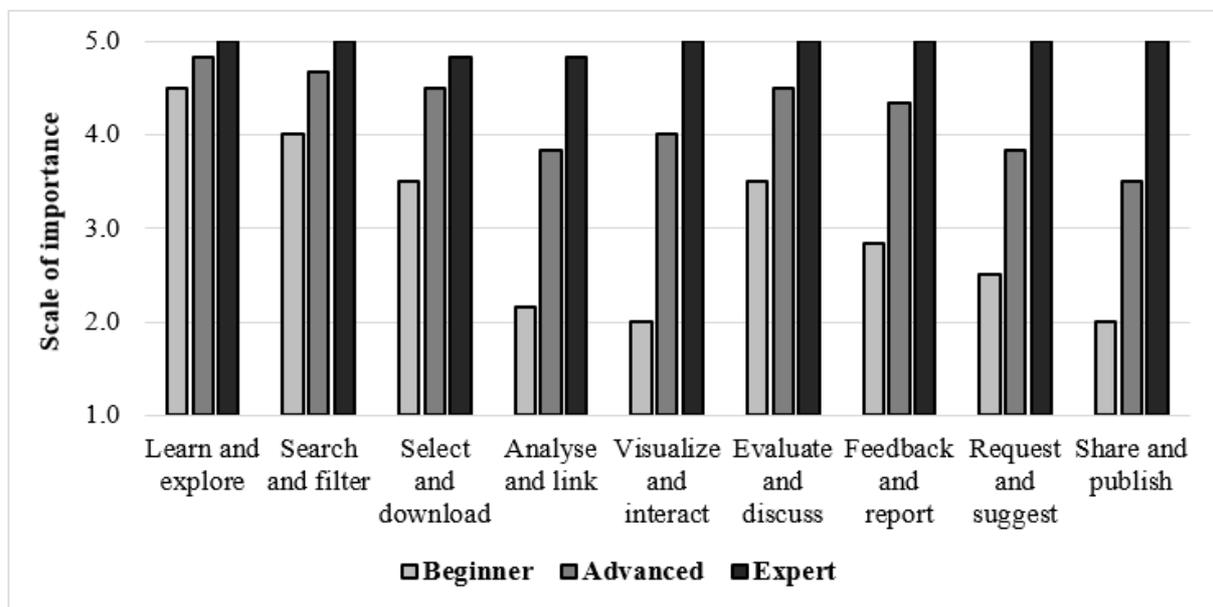

Figure 6. Suitability for different groups of users

Regarding our findings, recommendations towards development of stakeholders' knowledge and practice must be a priority for OGD to contribute to enhancing transparency. Data users need to have skills and knowledge to not only search through data sets, but they should be able to visualize, process, analyse, and assess potential impacts of these data. On the other hand, with respect to data users with different levels of knowledge and skills, the portals should provide features allowing that even a user with basic computer skills and literacy can be engaged and contribute to developing the OGD ecosystem. More precisely, the level of knowledge and skills cannot limit the use of these features. A detailed summary of the experts' comments who justified the evaluations of the feature-related transparency mechanisms are provided as follows. Regarding the RQ2, these findings should be used to design open data portals.

We identified **learn and explore** followed by **search and filter** as the most important categories of features for the users with basic skills. These should be presented on the homepage and help users to get information about OGD, where they came from, how they were collected or how the portal works. Gaining insights into how to use these data as well as features of the portal is a prerequisite for enriching users' knowledge and skills base that will lead to asking the questions, searching for solutions, exploring relevant data sets, engaging more stakeholders, and building new relationships towards increasing transparency. In the context of transparency, the portals should offer

as many of these features as possible, because when a user does not know what, where, and how to use it, no one will use it. The key features for the **learn and explore** category are:

- High-quality of tutorials, guidelines, and studies helping data users in learning how to use the portal and understand the source and limitations of these data.

- Dataset-related information such as metadata, data dictionary, or statistics on data sets and uses of the portal to identify specific data sets and actions of stakeholders.

- A list of current events, hackathons, and announcements as well as interviews relevant to the topic of open data and transparency.

- Examples (projects) of users leveraging open data, i.e., how open data can create impact including location, financing, jobs, data sets used, etc.

- A feature that enables to switch between the beginner and expert level documentation.

Dealing with **searching and filtering** capabilities, the portal should be designed to easily and quickly help data users find the right data set, including how to access data sets on the portal, and then filtering results according to various criteria. For this purpose, these features should be provided:

- Topics, and thematic categories together with relevant tags and keywords, including graphics, icons, and links for all data sets in order to enable data users effectively and efficiently search, filter, sort, and compare them.

- Filter options according to thematic categories, tags and keywords, organizations (publishers), administration level, data formats, resource type (data set, guide, terminology, web service, etc.), date attributes (period, update frequency), and open data licenses.

- Search results are returned in relevance order, but user interface should allow to order them by popular (views), rating, date added, last modified, name ascending and name descending.

Selecting and downloading a data set in the most suitable format and range is closely related to the context in which it will be analysed. Analysing and linking open data sets for transparency purposes include activities that require knowledge and understanding of the topic together with data analytics methods and techniques. The need of relevant skills is in this case limiting for most of the data users. In order to help users to select the most suitable data set and download it, these features should be available:

- Lists of selected / found data sets should include metadata to clearly describe each data set and help users in the selection process, an openness score according to a 5-star deployment scheme for open data makes the selection easier.

- Browsing through statistics (displayed in a chart) or dashboard to obtain information about data sets including number of visits, number of data sets, number of reuses, number of APIs, number of downloads, data sets by department or agency, data sets by thematic category, percentage of distributions by format, etc.

- Machine-readable formats to enable users to process data sets in a structured way, i.e., in CVS, JSON and RDF formats.

- Features to support automatic translation (export) into a wide variety of formats.

- The following options to download a data set: direct download through the URL available, a visit button that redirects a user, a download button for a single file according to a format chosen, and a download button for an archive file.

Features for **analysing and linking** should help users to analyse more data at a time and allow them to mash-up and link them for more effective analysis. However, analytical tools integrated in these portals may not be able to satisfy all the data users' requirements. These features are mostly designed for data users with expert skills and knowledge:

- API and Protocol and RDF Query Language (SPARQL) endpoints to enable users to access the resources from other systems, and options for more manual processing, such as CSV, JSON, RDF files, etc.
- Links to view a data set or its metadata within a browser in Hypertext Markup Language (HTML) or eXtensible Markup Language (XML) format to facilitate further data processing.
- An easy to interpret data set preview for each data set in machine-readable format, i.e., graphics and symbols, font types, colour contrasts, multilingual, etc.
- It can be implemented as a preview in a table provided by a portal or open a data set with external applications (sites), such as plotly, which makes it easy to analyse data together.

Although **data visualization** may help to understand the context and meaning of data better, it also requires more advanced skills to work with interactive graphs and statistics. For users with basic skills, appropriate presentation of data in a table or static graph format is recommended. Advanced and expert users may use features that display data sets in various graphs and charts including interaction features enabling to change a chart type, change a data field, filter according to fields or import a new data set – mostly shapefiles, CSV files, or web services. These features help data users to explore data sets more in depth and make the portal more useful. In addition, providing links to visualizations created by third parties such as a data set extent (map view) spatial coverage of a data set using OpenStreetMap or ArcGIS tools enables to view data in other contexts. Data users can also download data sets and the visualization can be done within their applications in off-line mode. This group of features also applies to various statistics and interactions between data publishers and users. In this regard, portals should contain lists of activities for each publisher to ensure transparency of their interactions with users, statistics about uses of a data set by other users, such as number of views and downloads, applications using the data set, etc. An initiatives' map in which all open data initiatives and portals on different administrative levels in particular country are displayed including a brief description is also important.

Enabling data users to **evaluate, comment and discuss** on data sets, including sharing them, are important steps forward for transparency. Supporting discussion forums on different topics to exchange ideas among the users is recommended as well as capabilities to comment on a data set and start a discussion using an account created in the portal or social media account. Connecting open data portals to social media and increasing visibility of data sets leads to developing new relationships and networks between stakeholders. Portals should contain capabilities allowing the collection of user ratings on a data set. A private space for registered users to collaborate should be also provided.

Providing **feedback and reporting** on data sets or features of the portal ensure that data users can be engaged in public discussions and decision-making. If this interaction is reciprocal it leads to their sense of responsibility towards public actions and policies. By cooperating with data users,

public authorities can get valuable information about what is needed to be improved, changed, or fixed, thereby increasing the efficiency of public service delivery. The feedback received should be shared with other public sector agencies and institutions to facilitate mutual learning and the transfer of experience between authorities (e.g., creating good practice examples / manuals, etc.) or linking data between different departments. Giving feedback and reporting features should enable:

- Data users to report a problem or raise a question on any other topic through a form.
- To distinguish between capabilities for individual data sets and for the portal itself, i.e., the feature to report a problem or error with a data set (a new issue or a solution for specific data that is currently on the portal) or the feature provided by the portal.
- To submit feedback on a data set from the users to publishers by using an account created in the portal, social media account, or without any registration required.
- To ask a question including a list of all questions and their status – all actions should be publicly available, including their status of completion.

Requesting and suggesting new data sets or data sources as well as providing a list of requests that were received from other users increase transparency and trust in public actions. Features dealing with requesting and suggesting should help users to request a new data set or other source, such as a document. Portals should provide:

- A form to request new data set that is not currently available on the portal and a list of requests that were received from users.
- The requests are public, searchable, include the current status of request processing (organization contacted, under review, in-progress, closed) and the outcome of the request (data set released, released in part, not collected, request rejected due to).
- Capabilities allowing the involvement in the active requests, i.e., express interest in the same data set, such as number of votes for a data set.
- A feature to check on its status and add their requests by clicking on thumbs up.
- Features dealing with uploading an application or other solution that reused open data (data sets published on the portal).
- A form to suggest new resource that is not currently available on the portal.
- A feature to submit a new story describing an application or solution using a specific data set on the portal, including how have open data made a difference, how have they increased transparency, informed citizens, supported better decision-making, etc.
- A feature to report the suspected corruption anonymously (found by analysing open data sets).

Public **sharing and publishing** of online information reduce the possibility of an unjustified advantage resulting from asymmetry of information since its existence could lead to corrupt practices, such as bribery to obtain information that is not freely available. Discussing public data (or budgets) provides the opportunity to exchange insights for citizens who are not indifferent to possible corrupt practices. These features should be available:

- To help users in sharing and publishing their own resources, i.e., modified data sets, documents, articles, etc.

- Lists of mobile as well as web-based applications created by reusing published data sets because they are more user-friendly and summarize and visualize selected data sets in more effective and informative ways than raw data sets.

- To submit a new application or upload and share other resources (a new data set, a modified data set, a reuse – use case of published article, computer graphics, etc.).

- To search and filter applications and other resources according to publishers, subject (thematic category), device formats (web, iOS, Android, etc.), date created (published) and last modified.

- A feature to create a distribution channel and share a link to a data set on social sites.

- A feature to provide private sharing with individuals or groups, so authorized users can access (and combine, export, and publish) all data in one place.

Finally, there are some notable observations resulting from the content analysis and the Delphi process that are helpful to understand each feature and its potential contribution to enhancing transparency through OGD reuse. In the first place, governments have a responsibility to promote open data portals, adoption of policies and best practices in order to ensure that OGD are available in one place for everyone to be reused. Information about what data are published and why some data cannot be released must be provided to establish public trust in government actions. Addressing the issues of security, privacy, confidentiality, protection of personal data and intellectual property must be a priority for governments. Improving the culture of openness and transparency also requires building awareness programs and campaigns, providing education and support to increase data literacy and skills needed to work with these data. Governments should establish relationships with other organizations working in the domain of transparency. They should stimulate cross-border cooperation in order to enable data interoperability and engage more data prosumers who will enrich the OGD ecosystem by creating and publishing their data sets. Monitoring, evaluating, and reporting on the progress and impact of OGD is also an important factor in increasing transparency.

These recommendations are primarily relevant for governments and local public sector agencies and institutions that should incorporate them into designing and developing their open data portals. Data users can also profit from them since they provide an overview of features that should be available to fully work with and reuse OGD. In addition, engagement of stakeholders is important to enhance transparency. Identification of feature-related transparency mechanisms and assignment of relative importance to each of them provides further context to this study and explains the need to consider stakeholders with different level of skills. We enrich current knowledge by this approach that clearly distinguished between characteristics OGD and mechanisms enabling to work with them. Most of the previous studies neglect this fact and fail to fully contribute to increasing of transparency.

5 Discussion and limitations

In this study, we have identified the categories of most common features and capabilities adopted by national open data portals worldwide regarding transparency-enhancing processes. There are three major contributions of this study resulting from the research questions. First, the features are directly gathered from the respective data portals and present the latest development of global OGD release and reuse. Second, the identified features are classified into categories forming the conceptual model that is more comprehensive and representative compared to previous studies. Third, the model serving as a benchmarking framework is evaluated using the Delphi method and the opinions of the experts are quantified to reflect different multidisciplinary views on this issue and present recommendations

towards developing open data portals that will better serve anti-corruption efforts of governments.

In line with other authors such as Bertot *et al.* (2010), Glassey (2017), Janssen *et al.* (2017), Klein *et al.* (2018), Lourenço (2015), Matheus and Janssen (2020), Safarov *et al.* (2017), or Saxena and Muhammad (2018), this study emphasizes the importance of transparency effects of OGD and open data portals. In addition, the identified features and capabilities complement and extend previous studies on this topic such as Alexopoulos *et al.* (2017), Bogdanović-Dinić *et al.* (2014), Chatfield and Reddick (2017), Máchová *et al.* (2018), or Thorsby *et al.* (2017). In order to effectively identify, disclose, and reuse OGD for transparency purposes, several conditions and requirements must be considered and embedded in the design of any open data portal. These can be found in Janssen *et al.* (2017). They proposed a concept of transparency-by-design according to which transparency should be considered as the main requirement for the development of the system, which must be met by the systems in order to ensure that the data that are made public create transparency. Our findings support the earlier studies by Jarke (2019) and Murillo (2015) who emphasised the need for government policies that focus on the utilization of new technologies focused on online data availability and the involvement of all stakeholders in public services. In this regard, the features and capabilities enabling engagement of various stakeholders in reusing OGD and contributing to improved transparency and accountability are the key ones.

However, despite the promising results obtained by the expert opinion poll, there are still limitations that need to be considered. The first limitation is represented by the selection of the expert opinion poll members. Although several stakeholders in different organizations and roles were interviewed, the sample may not have been fully representative. This issue was overcoming by evaluating the feature-related transparency mechanisms from the data user perspective in which different expertise and backgrounds come together. More detailed findings are obtained by this approach and enabling to differentiate between suitability for stakeholders with different levels of knowledge and skills. Second, the study conducted a qualitative investigation with a relatively small sample limited to the Central European countries only. It is expected that further research would conduct an international Delphi study with a larger sample size to identify and achieve consensus on priorities for this issue. We recommend examining whether identified feature-related transparency mechanisms are relevant in other countries. We are also going to validate our conclusions in real life conditions, i.e., investigating the readiness of public agencies and institutions to use open data portals as a tool to combat corruption.

Third, we conducted a study where the responses were gathered from the experts at a point of time and the comparison might already be outdated by the time that this paper is published. The OGD infrastructures are developing at a rapid pace. Fourth, our main investigation was limited to assessing the national open data portals but there are also portals on different administrative levels that may contain other features and capabilities. As reported by Zuiderwijk *et al.* (2013), open data portals and data infrastructures may have different purposes and different target groups, which probably require different features. They also search for different data sets. The question how to attract data users to reuse open data and which data sets could attract more users, so the impact of the portal can be high, is investigated through the concept of high-value datasets (Utamachant and Anutariya, 2018). Finally, further research will involve stakeholders in beginner, advanced and expert roles to explore to what extent are they able to use the identified features and capabilities of open data portals.

There can be also found challenges and obstacles decrease opportunities to achieve OGD's transparency effects (Matheus and Janssen, 2020). Recurrent drawbacks in open data portals were summarized by Reis *et al.* (2018), including recommendations on how data governance can tackle

much of the issues identified. Safarov *et al.* (2017) reported that most of the released public data sets on the OGD portals seem less relevant in terms of utilizing them for anti-corruption purposes. Thus, more research is needed to gain insights into how the data sets published on the open data portals are relevant in terms of utilizing them for anti-corruption purposes. Studies conducted by Ojo *et al.* (2016) and Saxena and Muhammad (2018) showed that the most common obstacle for using of open data portals is perceived the poor quality of open data provided on them. It is characterized by poor metadata, failure to present data appropriately to different audience and difficulty in locating data of interest. If these have no adequate quality, this barrier can result in discussions, confusions, less transparency and even in less trust in the government (Clarke and Margetts, 2014; Janssen *et al.*, 2017). According to Alexopoulos *et al.* (2017), the success of open data portals relies critically on three main characteristics of them: their information quality, their system quality, and their service quality. Regarding the issue of data providers and their role in ensuring the quality and access to relevant data, Kimball (2011) explored the way how public sector agencies and institutions develop open government training programs.

Other barriers are related to irrelevancy of data sets, the poor usability of portals and lack of examples of the prior uses of available data sets as well as applications built on top of OGD (Hogan *et al.*, 2017; Máchová *et al.*, 2018; Ruijer *et al.*, 2017). The development of OGD infrastructures by public sector agencies and institutions is also seen as a major issue for fostering OGD initiatives and disclosing OGD (Attard *et al.*, 2015). Other challenges include technical maintenance; licensing, confidentiality; release of easiest data only; difficulty in development of metrics and no guarantees for realizing value (Máchová and Lněnička, 2017; Matheus and Janssen, 2020).

6 Conclusion

Summing up the results, it can be concluded that the article contributes to the topic of anti-corruption efforts in a time when transparency reforms are increasingly based on OGD and open government movement. For this purpose, various open data portals are launched worldwide aiming to provide a platform with features and capabilities that should help to improve these efforts. Open data portals are a critical part of any data infrastructure since they connect data providers with data users in terms of online information and enabling to build relationships needed to increase transparency. Therefore, we identified the key features and capabilities enhancing transparency efforts in order to improve the design of these portals and facilitate the access to data sets published on them.

In this regard, open data portals should provide features that would serve as both a point for finding and accessing data sets and applications that reuse them and a public space where stakeholders are encouraged to ask questions, request and suggest data, publish their resources, and contribute to public debate. By promoting this active engagement and sharing of online information, governments can help accelerate the building of links among different stakeholders, especially public officials responsible for the supply of data and data users. If these key concepts and overall goals are well understood, accepted, and shared by all stakeholders, a new ecosystem that enhances transparency can be build. However, the ability of OGD to provide positive effects is limited to the willingness of government and its institutions to recognize their importance and adopt proactive initiatives. The design and development of open data portals is the critical step since it represents a central point for accessing OGD. Governments have also an opportunity to collaborate more with businesses and citizens in developing enhanced services, and to make effective use of new technologies such as cloud computing, Internet of Things, smart cities, and big data for greater transparency.

Although we found that the range of features and capabilities provided by open data portals is wide, there are duplicates and overlaps among them. Thus, we provided a classification of them into categories that represent the key feature-related transparency mechanisms. We also considered different skills of stakeholders and presented recommendations that can increase the number of active data users. For each one of the defined mechanisms, however, it is still necessary further research to evaluate whether these are concise, robust, comprehensive, and explanatory. For future research, ongoing studies on best practices in designing open data portals are needed to enhance transparency through reusing OGD. It will be explored whether the increased transparency (enabled through features and capabilities of open data sources) leads to better discovery of e-government services.

References

Afful-Dadzie, E., and Afful-Dadzie, A. (2017), "Open Government Data in Africa: A preference elicitation analysis of media practitioners", *Government Information Quarterly*, Vol. 34, No. 2, pp. 244-255.

Alexopoulos, C., Diamantopoulou, V., and Charalabidis, Y. (2017), "Tracking the evolution of OGD portals: A maturity model", in *International Conference on Electronic Government (EGOV 2017)*, Cham: Springer International Publishing, pp. 287-300.

Attard, J., Orlandi, F., and Auer, S. (2016), "Data driven governments: Creating value through open government data", in Hameurlain, A., Küng, J., Wagner, R., Anjomshoaa, A., Hung, P.C.K., Kalisch, D. and Sobolevsky, S. (Eds), *Transactions on Large-Scale Data- and Knowledge-Centered Systems XXVII. Lecture Notes in Computer Science, 9860*, Berlin Heidelberg, Springer, pp. 84-110.

Attard, J., Orlandi, F., Scerri, S., and Auer, S. (2015), "A systematic review of open government data initiatives", *Government Information Quarterly*, Vol. 32, No. 4, pp. 399-418.

Bertot, J.C., Jaeger, P.T., and Grimes, J.M. (2010), "Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies", *Government Information Quarterly*, Vol. 27, No. 3, pp. 264-271.

Bogdanović-Dinić, S., Veljković, N., and Stoimenov, L. (2014), "How open are public government data? An assessment of seven open data portals", in M.P. Rodríguez-Bolívar (Ed.), *Measuring E-government Efficiency, Public Administration and Information Technology*, New York, Springer, pp. 25-44.

Chatfield, A.T., and Reddick, C.G. (2017), "A longitudinal cross-sector analysis of open data portal service capability: The case of Australian local governments", *Government Information Quarterly*, Vol. 34, No. 2, pp. 231-243.

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*, Vol. 43, No. 3, pp. 483-502.

Clarke, A., and Margetts, H. (2014), "Governments and citizens getting to know each other? Open, closed, and big data in public management reform", *Policy & Internet*, Vol. 6, No. 4, pp. 393-417.

Glassey, O. (2017), "Proactive transparency and open data: A tentative analysis", *in International Conference on Electronic Government (EGOV 2017)*, Cham, Springer International Publishing, pp. 314-323.

Gliem, J.A., and Gliem, R.R. (2003), "Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for likert-type scales", in *2003 Midwest Research to Practice Conference in Adult, Continuing and Community Education*, Columbus, OH, Ohio State University.

Goel, R.K., Nelson, M.A., and Naretta, M.A. (2012), "The internet as an indicator of corruption awareness", *European Journal of Political Economy*, Vol. 28, No. 1, pp. 64-75.

Grisham, T. (2009), "The Delphi technique: a method for testing complex and multifaceted topics", *International Journal of Managing Projects in Business*, Vol. 2, No. 1, pp. 112-130.

Hogan, M., Ojo, A., Harney, O., Ruijer, E., Meijer, A., Andriessen, J., Pardijs, M., Boscolo, P., Palmisano, E., Satta, M., Groff, J., Baker, M., Détienne, F., Porwol, L., Scarano, V. and Malandrino D. (2017), "Governance, transparency and the collaborative design of open data collaboration platforms: Understanding barriers, options, and needs", in *Government 3.0–Next Generation Government Technology Infrastructure and Services*, Cham, Springer International Publishing, pp. 299-332.

Janssen, M., Matheus, R., Longo, J., and Weerakkody, V. (2017), "Transparency-by-design as a foundation for open government", *Transforming Government: People, Process and Policy*, Vol. 11, No. 1, pp. 2-8.

Jarke, J. (2019), "Open government for all? Co-creating digital public services for older adults through data walks", *Online Information Review*, Vol. 43, No. 6, pp. 1003-1020.

Kimball, M.B. (2011), "Mandated state-level open government training programs", *Government Information Quarterly*, Vol. 28, No. 4, pp. 474-483.

Klein, R.H., Klein, D.C.B., and Luciano, E.M. (2018), "Identification of mechanisms for the increase of transparency in open data portals: An analysis in the Brazilian context", *Cadernos EBAPE.BR*, Vol. 16, No. 4, pp. 692-715.

Lindstedt, C., and Naurin, D. (2010), "Transparency is not enough: Making transparency effective in reducing corruption", *International Political Science Review*, Vol. 31, No. 3, pp. 301-322.

Lněnička, M., and Komárková, J. (2019), "Big and open linked data analytics ecosystem: Theoretical background and essential elements", *Government Information Quarterly*, Vol. 36, No. 1, pp. 129-144.

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.

Lourenço, R.P., Piotrowski, S., and Ingrams, A. (2017), "Open data driven public accountability", *Transforming Government: People, Process and Policy*, Vol. 11, No. 1, pp. 42-57.

Máchová, R., Hub, M., and Lněnička, M. (2018), "Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective", *Aslib Journal of Information Management*, Vol. 70, No. 3, pp. 252-268.

Máchová, R., and Lněnička, M. (2017), "Evaluating the quality of open data portals on the national level", *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 12, No. 1, pp. 21-41.

Matheus, R., & Janssen, M. (2020), "A systematic literature study to unravel transparency enabled by open government data: the window theory", *Public Performance & Management Review*, Vol. 43, No. 3, pp. 503–534.

Meijer, A. (2009), "Understanding modern transparency", *International Review of Administrative Sciences*, Vol. 75, No. 2, pp. 255-269.

Moon, M. J. (2020), "Shifting from old open government to new open government: Four critical dimensions and case illustrations", *Public Performance & Management Review*, Vol. 43, No. 3, pp. 535-559.

Murillo, M.J. (2015), "Evaluating the role of online data availability: The case of economic and institutional transparency in sixteen Latin American nations", *International Political Science Review*, Vol. 36, No. 1, pp. 42-59.

Nikiforova, A., & McBride, K. (2021), "Open government data portal usability: A user-centred usability analysis of 41 open government data portals", *Telematics and Informatics*, Vol. 58, 101539.

Noveck, B.S. (2017), "Rights-based and tech-driven: Open Data, freedom of information, and the future of government transparency", *Yale Human Rights and Development Law Journal*, Vol. 19, No. 1, pp. 1-46.

Ojo, A., Porwol, L., Waqar, M., Stasiewicz, A., Osagie, E., Hogan, M., Harney, O. and Zeleti, F.A. (2016), "Realizing the innovation potentials from open data: Stakeholders' perspectives on the desired affordances of open data environment", in *Working Conference on Virtual Enterprises*, Cham, Springer International Publishing, pp. 48-59.

Okoli, C., and Pawlowski, S.D. (2004), "The Delphi method as a research tool: An example, design considerations and applications", *Information & Management*, Vol. 42, No. 1, pp. 15-29.

Reis, J.R., Viterbo, J., and Bernardini, F. (2018), "A rationale for data governance as an approach to tackle recurrent drawbacks in open data portals", in *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, New York, ACM, p. 73.

Ruijer, E., Grimmelikhuijsen, S., Hogan, M., Enzerink, S., Ojo, A., and Meijer, A. (2017), "Connecting societal issues, users and data. Scenario-based design of open data platforms", *Government Information Quarterly*, Vol. 34, No. 3, pp. 470-480.

Safarov, I., Meijer, A., and Grimmelikhuijsen, S. (2017), "Utilization of open government data: A systematic literature review of types, conditions, effects and users", *Information Polity*, Vol. 22, No. 1, pp. 1-24.

Saxena, S., and Muhammad, I. (2018), "The impact of open government data on accountability and transparency", *Journal of Economic and Administrative Sciences*, Vol. 34, No. 3, pp. 204-216.

Sayogo, D.S., Pardo, T.A., and Cook, M. (2014), "A framework for benchmarking open government data efforts", in *Proceedings of the 47th Hawaii International Conference on System Sciences*, IEEE, pp. 1896-1905.

Thorsby, J., Stowers, G.N., Wolslegel, K., and Tumbuan, E. (2017), "Understanding the content and features of open data portals in American cities", *Government Information Quarterly*, Vol. 34, No. 1, pp. 53-61.

Utamachant, P., and Anutariya, C. (2018), "An analysis of high-value datasets: a case study of Thailand's open government data", in *2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, IEEE, pp. 1-6.

Veljković, N., Bogdanović-Dinić, S., and Stoimenov, L. (2014), "Benchmarking open government: An open data perspective", *Government Information Quarterly*, Vol. 31, No. 2, pp. 278-290.

Welle Donker, F., and 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.

Wirtz, B. W., and Birkmeyer, S. (2015), "Open government: Origin, development, and conceptual perspectives", *International Journal of Public Administration*, Vol. 38, No. 5, pp. 381-396.

Zuiderwijk, A., Janssen, M., and Parnia, A. (2013), "The complementarity of open data infrastructures: An analysis of functionalities", in *Proceedings of the 14th Annual International Conference on Digital Government Research*, New York, ACM, pp. 166-171.