

# Evaluating the Quality of Open Data Portals on the National Level

Renata Máchová<sup>1</sup> and Martin Lněnička<sup>1,2</sup>

<sup>1</sup> University of Pardubice, Faculty of Economics and Administration, Pardubice, Czech Republic,  
<sup>1</sup>renata.machova@upce.cz, <sup>2</sup> martin.lnenicka@student.upce.cz

Received 30 October 2015; received in revised form 11 May 2016; accepted 17 May 2016

## Abstract

Over the last few years, governments worldwide have started to develop and implement open data initiatives to enable the release of government data in open and reusable formats without restriction or charge for their use by society. As a result, a large number of open data repositories, catalogues and portals have been emerging in the world. The efficient development of open data portals makes it necessary to evaluate their quality systematic, in order to understand them better and assess the various types of value they generate. Citizens also expect data disclosed by official authorities to have quality in the sense that they are official data and therefore should be accurate and reliable. Consequently, the aim of this paper is to examine and compare the quality of these portals. For this purpose, a benchmarking framework is proposed and validated to evaluate the quality of open data portals on the national level. The results obtained show that the number of datasets online and the sophistication of open data portals and their functions differ, reflecting the lack of harmonization and the need for quality standards. In particular, the United Kingdom, India and the United States have published many datasets and launched advanced portals.

**Keywords:** Open data, Open government, National open data portals, Benchmarking framework, Content analysis, Evaluation, Quality

## 1 Introduction

The world economy has become a data-centric one and as a result, those with capabilities to extract maximum benefits from their data will have the power at political, social, cultural and, especially, economic level [5]. Therefore, over the last few years, an increasing number of governments have started to open up their data. This so-called Open government movement has resulted in the launch of numerous open data portals and infrastructures that aim at providing a single point of access for government data and explore their impacts [24], [30], [53], [58]. By publishing these data on open data portals, the governments are giving it back to the citizens, which indirectly paid for their creation with their taxes in the first place [25]. The emergence of these portals changes the way both citizens and researchers look for accountability-related data [32]. The main point here is to create a fast change towards data sharing and transparency. This change could be helpful to gain the benefits from opening up further data silos in the public sector [55].

Although there are many different sources of data, government data are particularly important because of their scale, breadth, and status as the canonical source of information on a wide range of subjects [50]. They are called Open Government Data (OGD). It is important to explicitly mention at this point that OGD is not an equivalent to, but a subcategory or subset of, open data, which may equally originate in the commercial, academic or third sectors [18]. Because of large amounts of data produced by the public sector, the open data model has evolved into the open big data model [11], [31], [38]. This intersection of open and big data is mostly about integrating multiple data sources, e.g. on the international, national, regional and local of the public sector and international organizations [31]. Marton et al. [38] stated that the fundamental concepts of open and big data are technical in nature as they were developed in the fields of computer science and engineering. They are both gathered for a purpose and then normally repurposed [11]. While big data are mostly characterized as large in volume, gathered at high-speed and may be also unstructured and come from many sources, open data are about standards on how to make data machine-readable, and hence linkable [38].

Governments and other public sector authorities generate and collect vast quantities of data through their everyday activities, such as managing pensions and allowance payments, tax collection, recording traffic data and issuing official documents [5], [55]. For example, these data are the largest single source of information in Europe with an estimated market value of 32 billion Euros [16]. Buchholtz et al. [5] then estimated that aggregate direct and indirect economic impacts from the use of open big data across the whole European Union (EU) economy are of the order of billions Euros annually. Hence, these data have a significant potential for re-use for developing new products and services, possibly in creative combinations with other data sources [8]. Zillner et al. [55] estimated overall economic gains of 40 billion Euros a year in the EU; besides the economic benefit as such, the EU recognizes additional value like the contribution to address societal and industrial challenges, achieve efficiency gains through sharing data inside and between public administrations and foster participation of citizens in political and social life and increasing transparency of government. The possibilities to better use these available data are growing due to the technical facilities and advancement to merge and analyze different datasets [21]. Areas of interest related to the application of these technologies include data analytics for the advance analysis of large datasets (benefits for e.g. fraud detection and cyber security), improvements in effectiveness to provide internal transparency and improvements in efficiency for providing better personalized services to citizens [30], [54], [55]. During the last years, a vast number of open data communities that develop new ideas and applications, have emerged around OGD portals [42]. Therefore, open big data should be the goal where possible. The only, minimum and non-radical, demand is: openness by default. The sooner governments are opening up their data the higher the returns [55]. However, disclosing these huge amounts of data does not necessarily equate to more transparency and does not necessarily facilitate accountability [32]. Except the economic importance, there are additional issues concerning the regulation of government data such as discoverability, harvesting, community engagement and interoperability [20], [50]. Also the quality of these data and related portals may vary from country to country, which can affect their value for users.

An ability to discover the relevant data is a prerequisite to unlocking the potential of open data. Creating a portal of available datasets is a one way how to make these datasets more accessible and easier to find [27]. On the other hand, with the different data management systems of open data portals and related open data initiatives, there is a great diversity in their content, functionality and technology standards [4], [30]. But most importantly, they vary in their usefulness and suitability to their task [4]. The extraction of valuable information coming from these different data sources then requires the quality evaluation of them [6]. Calero et al. [7] addressed the evaluation of the quality of a web portal by defining a data quality model containing 33 data quality attributes grouped into four data quality categories. While some aspects of open data quality align with the ones of web portals, domain-specific quality perspectives in the context of open data (e.g. data management system, the openness of provided data based on the license or format, metadata) need to be identified and evaluated. The quality of data plays an essential role in the use of open data portals and a certain level of data quality is critical for OGD use [57]. However, Umbrich et al. [48] argue that despite the enthusiasm caused by the availability of a steadily increasing amount of openly available data, first critical voices appear addressing the emerging issue of low quality in the metadata and data source of open data portals, which is a serious risk that could disrupt the open data project. Kučera et al. [27], Umbrich et al. [48] as well as Zuiderwijk and Janssen [57] then claim that there is a need for a quality evaluation and benchmarking framework to better understand quality issues in open data portals and study the impact of improvement methods over time.

Therefore, this paper contributes to filling this research gap, by conducting a literature review to propose a new benchmarking framework focusing on the quality evaluation of various dimensions of these portals and presenting an analysis of currently available portals on the national level in the world, in particular, to understand whether these portals provide data in a way that in fact facilitates their use, re-use and redistribute public accountability.

The organization of the paper is as follows. In Section 2, the research methodology is formulated and the research method is chosen and described. Section 3 introduces the literature review and background on open data and related initiatives, benefits, risks and impacts of opening government data, open data portals, their quality and evaluation requirements, etc. Section 4 presents the results of open data portals quality evaluation. In Section 5, these findings are discussed. Finally, concluding remarks are provided in Section 6.

## 2 Research Methodology and Methods

The research method is adopted for conducting a multi-dimension analysis [39], which clearly distinguishes between dimensions and their scores, and the quality evaluation of the national open data portals in the world. It consists of four distinct stages:

1. Identification of OGD sources as open data portals on the national level as were defined in [30] and ranks represented by the Global Open Data Index (GODI) [40] and also the Open Data Barometer (ODB) Global Report and index [52];
2. Proposal of the benchmarking framework (definition of the quality dimensions, and the particular metrics for each of them) based on the systematic literature review, as defined by Petticrew and Roberts [41], i.e.: A review that aims to comprehensively identify all relevant studies to answer a particular question, and assesses the validity (or soundness) of each study taking this into account when reaching conclusions. Therefore, the following steps are defined:
  - Define search terms and keywords search strategies based on the defined research question;
  - Select sources (digital libraries) on which to perform search;
  - Application of search terms and keywords on sources;
  - Assess the validity of studies identified in the search; and
  - Selection of primary studies by application of inclusion and exclusion criteria on search results.
3. Data collection with the use of questionnaires method as research data sampling technique; and
4. Data processing and results presentation, which include calculation of various descriptive statistics, such as frequencies and relative frequencies of all values for each of the metrics, and construction of various charts, using the Microsoft Excel software.

The study is based on quantitative techniques according to the relevant literature by Maylor and Blackmon [39]. More precisely, it follows their recommended analysis approach based on the data collection method and the dimensions used. It starts with the research questions that can be answered by collected data. The next step is the definition of the process for collecting and analyzing data. The final step is looking for trends and patterns, i.e. interpreting data.

## 3 Literature Review and Background

In addressing the literature review and background, multiple research streams that are associated with this study will be examined. The first section will define open data and related initiatives. It will be followed by benefits, risks and impacts of opening government data. After that, the issue of open data portals, their description and classification will be presented and discussed. It will be followed by open data portals quality and evaluation requirements. Finally, the literature related to benchmarks and models for evaluating the progress of open data portals will be examined.

### 3.1 Open Data and Related Initiatives

The topic of open data is generating interest among practitioners in the public sector as well as in the private sector [20] and continues to grow driven in part by pressure for increased public sector transparency and in part by the current enthusiasm for big data and data analytics [9]. Open government acts as an umbrella term for many different ideas and concepts. Open data ecosystems are often government ecosystems, as much open data are published by governments, although elements for these ecosystems can also be provided by the private sector. Open government

ecosystems then emphasize the multiple and varying interrelationships between data, open data providers, open data users, material infrastructures and institutions [58]. Mostly, they are equated with e-government and the usage of modern Information and Communication Technologies (ICT) and related tools [5], [15], [43], [47].

However, this topic is not entirely new, as the re-use of Public Sector Information (PSI) has been the subject of EU directives [16]. The first PSI Directive was adopted at the end of 2003. It introduces a common legislative framework regulating how public sector institutions should make their information available for re-use in order to remove barriers such as discriminatory practices and monopoly markets by harmonizing the regime for the re-use of PSI [16], [43]. Then it was revised in 2009 and again in 2013, bringing more public institutions within scope and encouraging free or marginal-cost, rather than recovery-cost, pricing: reflecting what was by then already practice in many EU states. Open by default is becoming one of the foundational principles of open data-related pieces of legislation related to this EU Directive on PSI [50]. Open data initiatives are a starting point for boosting a data market that can take advantage from open information (content) and the big data technologies. Therefore, active policies in the area of open data can benefit the private sector, and in return facilitate the growth of this industry. At the end this will benefit public budgets with an increase of tax incomes from this growing data industry [55]. The number of open data initiatives has grown from two to over three hundred in the period 2009-2014 [21], [43], and the membership in the Open Government Partnership (OGP) has gone from eight in 2011 to sixty-nine participating countries in 2016.

While the implementation of open data initiatives seem similar across governments (i.e., mostly through a centralized web portal, where datasets can be downloaded by the public), governments may have different motivations for embarking on open data initiatives [20], [24]. Recent studies have also shown that current initiatives employ different approaches for providing data and exhibit important limitations such as data duplication [23], [24]. Therefore, Sayogo et al. [44] conducted an in-depth evaluation of selected cases to justify the application of a proposed framework for understanding the status of OGD initiatives. Reggi and Ricci [43] explored the information-based strategies that EU Regions and Member States are implementing when publishing government data on the Web. Cohesion Policy and its Structural Funds, which involve all EU Regions and Member States, are the ideal context to verify the presence of different approaches to the publication of government data. They identified three approaches: user centered, which shows the effort to make these data understandable by non-technically oriented citizen, clearly represented, and accessible to users; stewardship, which is defined by many desirable characteristics aimed at assuring accuracy, validity, security, management, and preservation of information holdings; and re-user centered, which emphasized the importance to download the data in a machine-readable format and other characteristics related to data quality. Lee et al. [29] then examined open data initiatives in the some of the world's most innovative countries and noted that they can either be government-led or community-led.

Open data are a piece of content or data if anyone is free to use, re-use, and also redistribute them – subject only, at most, to the requirement to attribute and share-alike. Most of open data are actually in raw form. However, republishing does imply citing the original source not only to give credit but to ensure that these data have not been modified or misrepresented [19], [20], [26]. Linked data then describe a method of publishing structured data so that it can be interlinked and become more useful through Semantic Web technologies such as Uniform Resource Identifiers (URI), Resource Description Framework (RDF), vocabularies and ontologies. Linked data are a way of publishing data in such a way that they can facilitate the interaction between different data sources or developing advanced value-added e-services by combining different datasets from multiple OGD sources; also, the value of any kind of data increases each time they are being re-used and linked to another source, and this can be facilitated and triggered by providing informative and explanatory data about each available dataset, i.e., metadata [14], [15], [17]. These ideas then gave rise to the development of Linked Open Data (LOD), which are the combination of both, to structure data and to make them available for others to be re-used without any restrictions. Data interlinking practice is highly recommended for lowering technological and cost barriers of data aggregation processes [15], [17], [50].

Use and re-use of these data means using them in new ways by adding value to them, combining information from different data sources, making mash-ups and new applications, both for commercial and non-commercial purposes [13], [15], [30]. The core idea behind OGD is just very simple: government data should be a shared resource. Making data open is valuable not only for the government departments that collect and release these data, but also for citizens, businesses and other parts of the public sector, because OGD has limited value if these data published are not utilized, which means involving stakeholders and focusing on developing sustainable ecosystems of users [17], [52]. Basically, in an open data network there could be cooperation of various stakeholders to facilitate the use of OGD. Also, there could be competition between businesses using open data, for example, to obtain open data end-users as (paying) customers for services that they have developed based on OGD. There could also be competition between open data providers, since they may want to promote their organization by stating that they open larger amounts of data or more datasets than other public sector institutions do [29], [54], [58].

### 3.2 Benefits, Risks and Impacts of Opening Government Data

Both Janssen et al. [20] and Ubaldi [47] provided a comprehensive discussion of the challenges in the OGD domain. Kucera and Chlapek [26] presented a set of benefits that can be achieved by publishing OGD and a set of risks that should be assessed when a dataset is considered for opening up. Also Cowan et al. [11] used several practical examples in an attempt to illustrate many of the related issues and allied opportunities of open data. Furthermore, different authors have confirmed that releasing government data in open formats creates considerable benefits for

citizens, businesses, researchers, and other stakeholders to understand public or private problems in new ways through advanced data analytics [20], [56]. For end-users and society in general, open data will help to obtain and integrate required information more efficiently and successfully manage the transition towards a knowledge-based economy and information society [5], [50]. However, Yang and Kankanhalli [54] stated that despite public institutions actively promoting the use of their data by organizing events such as various challenge competitions, the response from external stakeholders to leverage OGD for innovative activities has still been lacking. Also, the findings of Zuiderwijk and Janssen [56] are in agreement with the claim that results of data re-use are not discussed and only little feedback is gained by data providers (public sector). Therefore, public sector institutions must have processes in place clearly defining which data to share with the users in which formats, at what time intervals, and under which licenses, ensuring no restrictions on re-use of these data [49].

Ideally, making these data available on the Web would lead to more transparency, participation and innovation throughout society. However, just publishing the data on the Web is not enough. To truly advance the open society, the publication platforms need to fulfill certain legal, administrative as well as technical requirements [4]. In practice, gaining access to raw data, placing them into a meaningful context, processing and extracting valuable information from them is often extremely difficult. As a result, during the last couple of years different solutions have been developed to support the whole lifecycle of these data re-use, i.e., data discovery, cleaning, integration, processing and visualization [17], [23]. According to [56], open data process consists of all activities between the moment that data are starting to be created and the moment that data are being discussed, including the activities to publish, find and re-use them. At least open data publishers and users are involved, but often many more stakeholders are involved, such as open data facilitators, brokers (organizations that bring together open data users and producers), citizens, businesses or open data legislators (the EU and national political parties) [21], [56].

Jetzek et al. [21] developed a conceptual model portraying how open data as a resource can be transformed to value. They showed the causal relationships between four contextual, enabling factors (openness, data governance, capabilities and technical connectivity), four types of value generation mechanisms (efficiency, innovation, transparency and participation) and value. They concluded that the value generation mechanisms are dependent on the enabling factors. If this relationship is well understood, it is easier to choose the right datasets, data platforms and governance procedures. In 2014, they updated this model focusing on generating sustainable value from open data in a sharing society. They introduced these new value generating mechanisms [22]: information transparency, collective impact, data-driven efficiency and data-driven innovation. All the mechanisms are dependent on the private and public sector, together providing the motivation, opportunity and ability to generate value from data. They also claim that the motivation, opportunity and ability of individuals to use data for value generation are influenced by [22]: the incentives provided; the level of technical and legal openness of data; the maturity of data governance; the general data-related capabilities in society; and the technological maturity and prevalence. These models were later used and improved by Bílková et al. [3] and Máchová and Lněnička [34] to evaluate the impacts of open data in the economic, educational, environmental, health, politics and legislation, social, and trade and business development. Maier-Rabler and Huber [36] discussed the impact of open data on the relations among citizens, public sector, and political authority to engage them in collaboration, co-decision, co-development and shared responsibilities, while Geiger and von Lucke [15] analyzed the added value of freely-accessible government data and discussed challenges of OGD for public sector at the different administration levels.

Through the last 10-15 years, various e-government development frameworks and indices have been introduced to help assess the opportunities and challenges of e-government and related open government initiatives. The early 2010s has added new indices to the e-government research, which are focusing on the evaluation of open data impacts. These are, e.g., the Web Index and the ODB index produced by the World Wide Web Foundation (W3F), Open Knowledge Foundation's (OKF) GODI, the OURdata (Open, Useful, Reusable Government Data) Index introduced by the Organization for Economic Co-operation and Development (OECD) and the Public Sector Information (PSI) Scoreboard (PSIS) by the EU [35]. These frameworks also have influence on the proposal of a new benchmarking framework.

### 3.3 Open Data Portals, their Description and Classification

One of the key issues in adopting open government is the accessibility of open data, which are generally collectively provided in open data portals [53]. Therefore, one of the first problems to be solved when working with any data is where to find them. In using data, user needs exactly the right dataset, i.e., with the right variables, for the right year, the right category, etc. [50]. According to the definition, open data have to be well described and in a good quality for others to transform them into knowledge and make them useful [10]. In the last few years, an increasing number of governments have launched open data portals, specialized websites where a publishing interface allows datasets to be uploaded and equipped with high-quality metadata [47], [50]. The open data portal is a web-based system used to collect existing datasets from multiple sources that may be in different formats, and publish them on user-friendly dashboards that users may view, download and access via an Application Programming Interface (API). With user-defined tags, these datasets are organized into a searchable catalog [25], [50]. It is operated by a catalogue operator, which could be a government agency, citizen initiative, etc. Each portal offers different datasets that directly reflect data availability to public disclosure [27]. The actual dataset is not considered part of the catalogue record, but the catalogue record usually contains a download link or web page link from where the actual dataset can be obtained [12]. Each dataset can comprise several data sources [50].

Open data portals categorize open datasets according to their domains, providers, formats, and other properties for better accessibility of the data [53]. Open data portals also usually feature keyword search and various browsing interfaces to help users find relevant datasets and retrieve corresponding metadata to describe the institution releasing the dataset as well as content of dataset in addition to geography, jurisdiction and time period of data [51]. Dataset format also needs an immediate attention as it may lead to lot of issues of interoperability and integration [51]. Other issues that are related to the context of the dataset concern completeness and exhaustiveness, the representation of open data, the validity, the reliability, the clearness and comprehensiveness and the provision of reports about analysis of these data. In line with these content related issues, the overall data quality should be taken into account [13], [16], [19]. Data standards, codes, vocabularies and schemas are also important aspects of datasets [51]. As an alternative to making raw data directly available for download, several portals offer web-based data APIs that enable developers to access data within their applications [28]. A sufficient description of the portal should clearly distinguish themes from keywords. While themes are always chosen from a controlled vocabulary, tags are not [33]. Furthermore, a machine-readable file format is important to support automatic tools. Features regarding validity, quality and granularity are required to support a wide range of use cases and enable analysis results of high quality [4].

Not too long ago, metadata were only a concern of information professionals engaging in cataloging, classification and indexing. However, nowadays, there are many more creators and consumers of digital content which also needs to be cataloged [17]. Without sufficient metadata, such as descriptions or tags, neither manual nor automatic search can find the dataset and it will not be helpful for any user [4]. Metadata structure of the data portal summarizes common properties used to describe each dataset across the selected portal. It mainly includes attributes such as the dataset's name, description and the URL of the actual sources, i.e., files or service end points. Using these metadata, users can quickly find datasets they need with searching and filtering features [50]. In terms of metadata semantics, the most important initiative that a data portal should accommodate to facilitate interoperability is a RDF vocabulary named Data Catalogue Vocabulary (DCAT) by the World Wide Web Consortium (W3C). By using DCAT to describe datasets, publishers increase discoverability and enable applications easily to consume metadata from multiple catalogues [12], [33]. Some authors also proposed their own DCAT RDF vocabulary as an interchange format to enable standardized description of data catalogues, such as Maali et al. in [33].

Based on the geographical coverage (administration level), open data portal can be divided into the following groups [27]: local, which is owned by cities/towns or with only city/town coverage; regional, which is owned by a regional authority (county government or federal state government) or with regional coverage; national, which is owned by a central government institution or with nationwide coverage; and international, which is owned by an international institution or with the international coverage. Lněnička and Máchová [31] then extended this classification by adding a new level of open data aggregators as a basic and the most important category of data catalogs [19], [31]. Another categorization of open data sources can be made based on the web paradigm, i.e., they are based on the traditional Web 1.0 paradigm or the more recent Web 2.0 paradigm [1]. Based on the maturity of open data portals, Colpaert et al. [10] proposed a five stages system to represent the main function or affordance that the data portal is built or used for. These stages are ordered by the investment of time needed to be able to fully implement the stage. The categorization starts with portals linking to various datasets and continues towards a metadata portal for both the datasets and the re-use of the datasets. The fourth category takes care of the data publication itself. Finally, a data hub is set up where data become a common resource.

The open data portal is one of the solutions that should be used to significantly improve discoverability of free available datasets [27]. However, many governments focus on the development of a national OGD portal as if it was a higher priority than developing technical infrastructures to open up government data for others to use [47]. Infrastructures may improve the use of OGD by providing insight in how individuals can participate in data re-use and in the quality of open data [57]. Understanding the preconditions for effective OGD in a specific context is essential to set up websites that enable value creation, and lies at the core of the government data publishing responsibility. Much of current criticism on national OGD portals is based on the fact that governmental interest appears to be on presenting data in a particular fashion, which distracts from, and thereby limits, the increasing provision to users of data that they are really interested in using for their own purposes [47]. Janssen et al. [20] surfaced several factors inhibiting public use of open data such as the lack of explanation of the meaning of data, and the lack of knowledge to make sense of data. Martin et al. [37] presented seven categories of risks associated with open data portals: governance, economic issues, licenses and legal frameworks, data characteristics, metadata, access, and skills. Also, various factors from institutional to technical seem to affect the development and implementation of the OGD portals at the national level [20]. Therefore, it is sensible to argue that different nations have different capabilities in developing and implementing their OGD efforts [44].

### 3.4 Open Data Portals Quality and Evaluation Requirements

As the amount and the variety of data sources are increasing, it is important to create good metadata (descriptions, geographical coverage, limitations, etc.) in order to allow stakeholders, who may not be domain experts, to easily search and consume data. In a sense, the notion that all data disclosed should have quality or be intrinsically good is self-evident. But such a concept is not easy to pinpoint in the context of open government, and the requirement for data quality may be considered as encompassing several characteristics [32]. The risk of low (meta)data quality affects the discovery and consumption of a dataset in a single portal and across portals. On the one hand, missing

metadata directly affects the search and discovery services to locate relevant and related datasets for particular user needs. On the other hand, incorrect descriptions of the datasets pose several challenges for their processing and integration with other datasets [48].

The literature dealing with data quality provides different classifications of the data quality attributes, depending upon the perspective of the authors and the context tackled. On the other hand, in order to assess data quality a growing tendency towards considering the users' point of view exists. In fact, the most common definition of data quality is data that are fit-for-use, i.e. the ability of a collection of data to meet user requirements [7]. Data quality is usually described in the literature by a series of quality dimensions that represent a set of consistency properties for a data artefact [2], [6]. Batini et al. [2] published a detailed and systematic description of methodologies to assess and improve data quality. Methodologies are compared along several dimensions, including the methodological phases and steps, the strategies and techniques, the data quality dimensions, the types of data, and, finally, the types of information systems addressed by each methodology. Caballero et al. [6] focused on the evaluation of data quality. They claimed that more than ever the need for assessing the quality-in-use of datasets gains importance since the real contribution of a dataset to a value creation can be only estimated in its context of use. The most important characteristic for assessing the level of quality in use of heterogeneous datasets is consistency, which is divided into three parts: contextual, temporal and operational. Further evidence supporting the importance of the data quality can be found e.g. in Tien [46].

Measures on data quality can be applied in the open data domain. Open data domain-specific data quality criteria and measurements emerged recently to evaluate the quality of open datasets as well as portals [53]. According to the literature analyzed, open data portals, in order to fulfill quality informational needs, should address the whole range of entity types and requirements possible in each country institutional arrangement [32]. Also the semantics and language that each open data portal is tied to, is one of the most common and inherent quality challenges [42]. Research on the quality of open data portals has confirmed that a first wave of these portals mainly provides basic functionalities for uploading and downloading data [1], [8]. Existing portals and infrastructures often lack opportunity for data users to participate in improving published datasets [1], [57]. Lourenço [32] defined these eight key characteristics for open data portals: quality, completeness, access and visibility, usability and comprehensibility, timeliness, value and usefulness, granularity and comparability. Although open data success strongly depends on the quality of released datasets, there is a wide variety in the quality of the released datasets and users may also be concerned about the quality of open data [37], [42], [57].

Recent experiences also show that the quality of the catalogue records might affect the ability of users to locate the data of their interest. Therefore, Kučera et al. [27] discussed some limitations associated with the quality of the catalogue records, e.g. the metadata in the catalogue records is insufficient or correctness of the provided metadata is not checked, and proposed relevant techniques for its improvement. Open data publishers and users are often not aware of each other's needs and activities. For instance, many open data providers are primarily focused on making data available and do not know which format is preferred by users and how the way that they publish data can stimulate the use of open data [56]. Lourenço [32] then identified a set of requirements that dataset in the data catalogue needs to fulfill in order to contribute to the transparency of public agencies and allow for the accountability of the public sector institutions. Such requirements concern the type of entities covered by the dataset, the type of information types provided, the information seeking strategies supported and some qualitative aspects of open data provided. Kučera et al. defined four requirements of dataset quality [27]: accuracy-all information in a catalog record should correspond to the data described by the record and all the catalog records in the catalog should be accurate; completeness- all mandatory attributes of the record should be filled in and all published open government dataset should be registered in the catalog but there should be no duplicate catalog records; consistency-same terms or concepts should be used to classify data of the same type or category, unknown or missing information should be handled in the same way across the whole catalog; and timeliness- all information in the catalog record should be up-to-date and all the catalog records should be up-to-date.

In order for users to assess data quality, they need to understand the nature of the data and because data producers cannot anticipate all users, the provision of good quality metadata is as important as the quality of data themselves [55]. However, the role of public institutions in open data strategies and initiatives does not lie solely in the release of data. Other than increasing the variety and improving the quality of data made available to the public, there have been concurrent efforts by public institutions to motivate the use of open data for innovation activities by external stakeholders [40]. Kučera et al. [27] identified two types of strategies for improving data quality; namely data-driven and process driven. The first involves directly modifying the values of data, such as correcting invalid data values or normalizing data. The second involves the redesign of the data creation and modification processes in order to identify and correct the cause of quality issues, such as implementing a data validation step in the data acquisition process. Various efforts already exist to study different aspects of open data portals which are the main platforms to publish and consume datasets [48].

### 3.5 Benchmarks and Models for Evaluating the Progress of Open Data Portals

Since the launch of the first open data portals by the United States government in 2009 and the United Kingdom (UK) in 2010, an increasing number of countries have launched similar open data initiatives and data portals to make it easy for the public to find and use these data, which are available in a range of formats and span through a wide

range of domains [54]. This is in line with the findings of Umbrich et al. [47], which observed that the number of datasets and sources is continuously growing. Examples for the increasing popularity of data portals are OGD portals [20], data portals of international organizations and Non-Governmental Organization (NGO)s, scientific data portals as well as master data catalogues in large businesses [14], [19]. Numerous countries, including a good number of EU Member States, have followed along with some local (e.g. city) governments [50]. Many of these portals use Comprehensive Knowledge Archive Network (CKAN), a free, open-source data portal platform developed and maintained by Open Knowledge. As a result, they have a standard powerful API, which raises the possibility of combining their catalogues to create a single worldwide entry point for finding and using government data. Others then include Drupal Knowledge Archive Network (DKAN), Open Government Platform, Socrata, Prognoz or Junar. Similar to digital libraries, networks of such data catalogues can support the description, archiving and discovery of data on the Web [14], [50]. As well as the official public and private sector sponsored portals, there are numerous unofficial sources of open data, usually compiled by citizens, communities or aggregators.

Kalampokis et al. [24] revised existing e-government stage models and proposed an OGD stage model, which provides a roadmap for OGD re-use and enables evaluation of relevant initiatives' sophistication. Solar et al. [45] then proposed an open data maturity model to assess the commitment and capabilities of public institutions in pursuing the principles and practices of open data, which has a hierarchical structure consists of domains, sub-domains and critical variables. Alexopoulos et al. [1] developed a new model of the open data portal by extending its functionality using a wide set of capabilities for data processing, enhanced data modeling (flat, contextual, detailed metadata), commenting existing datasets and expressing needs for new datasets, datasets quality rating, users groups formation and extensive communication and collaboration within them, data linking, upload of new versions of existing datasets and advanced data visualization. Van der Waal et al. [50] described the key functionalities of open data portals and presented a conceptual model to make these portals the backbone of a distributed global data warehouse for the information society on the Web. Charalabidis et al. [8] presented and validated a methodology for evaluating these advanced second generation of ODG infrastructures and open data portals, which is based on the estimation of value models of them from users' ratings. They concluded that the highest priority should be given to the improvement of the data upload and data search-download capabilities, since they received low ratings from the users, and at the same time they have high impact on higher layers' value generation.

One of the first comparisons of the selected open data portals was conducted by Maali et al. [33] in 2010. They aimed to identify commonalities and overlap in the structure and to document challenges and practices. However, only seven data portals from five different countries were compared. Sayogo et al. [44] used web content analysis in order to demonstrate the application of data manipulation and engagement capability of open data portals from 35 countries. Verma and Gupta [51] compared 30 national (country) level data portals to find out the variety of formats in which different datasets are released. Their findings suggest that in general, open data portals development follows an incremental approach similar to those of e-government development stages. Furthermore, Umbrich et al. [48] continuously monitored and assessed the quality of 82 active open data portals, powered by CKAN across 35 different countries. For this purpose, they proposed six quality dimensions and related metrics: retrievability, usage, completeness, accuracy, openness and contactability. Their results then include findings about a steady growth of information, a high heterogeneity across these portals for various aspects and insights on openness, contactability and the availability of metadata. However, they assessed open data portals at the different administration levels.

Zuiderwijk and Janssen [57] then evaluated the usability of participation mechanisms and data quality indicators for open data portals with the use of six criteria, while Lourenço [32] assessed whether the current structure and organization of seven open government portals is adequate for supporting transparency for accountability. The author introduced a set of requirements identified based on the key characteristics of desired data disclosure from the literature on open government and transparency assessment. These requirements were used as a framework to analyze the structure and data organization of these portals. Yang et al. [53] compared categorization structures of open data portals by investigating the coherence, i.e. similarity, of the datasets in the same category. Braunschweig et al. [4] presented a survey of existing OGD platforms, focusing on their technical aspects. They studied over 50 portals operated by national, regional and communal governments, as well as international organizations, focusing on features such as standardization, discoverability and machine-readability of data.

Lněnička and Máchová [31] evaluated selected national open data portals. However, they compared only the EU Members States with the use of five criteria. Lněnička [30] later extended the list of portals to 67 countries, but with no further comparison. Petychakis et al. [42] analyzed the OGD sources developed in the EU27 from a functional, semantic and technical perspective, in terms of their thematic content, licensing, multilingualism, data acquisition, data discovery, data provision and data formats. They concluded that most of the datasets of the European OGD sources are published without a clearly defined or open license and about half of these OGD portals in their user interface support the native language of the corresponding country, while the other half are multilingual (they support one or more foreign languages as well).

However, none of the previously mentioned research papers evaluated the quality of open data portals and related datasets on the national level. Also, none of these research papers used more than 60 open data portals on a single administrative level. Therefore, a new benchmarking framework is proposed based on the previous literature review to solve these limitations. Only open data portals on the national level are considered to showcase the applicability of



the proposed framework. Furthermore, portals with fewer than 50 published datasets were excluded since it was considered that they would not have enough critical mass to provide useful insights, i.e., 67 countries are evaluated.

## 4 Results of Open Data Portals Quality Evaluation

Only open data portals on the national level are evaluated, no international, regional or local open data portals, and also no national statistical institutes or offices portals, which may also offer open data. The comparison is based on the rankings of the GODI and the ODB index from 2015, which evaluate the state of open data in selected countries in the world. Together, they cover 140 countries. The verification and validation process of the open data portal's existence consists of these steps: a keyword consisting of the name of the countries listed in the rankings mentioned above is inputted into general search engine Google together with *open data* or *open data portal*; the selected country is compared with the list available at other sources such as Site 1, Site 2 and Site 3; and the identified portal's URL is opened to examine whether it is in working condition. The portals, which were found at the beginning of 2015, are presented in [30]. Later that year, a new searching process was established, and 24 additional open data portals on the national level were found. It increases the total number of open data portals to 91.

### 4.1 Evaluating the Quality of Open Data Portals through API

As the first step in the quality evaluation of open data portals, a content analysis is used to compare selected open data portals on the national level. For this purpose, the API is used to access these portals. Therefore, only the portals, where the API is available, will be evaluated. Most of these 52 open data portals are powered by CKAN or DKAN. Postman Representational State Transfer (REST) client is used as the main tool to get JavaScript Object Notation (JSON) formatted results.

The first group of functions is focused on the lists of a portal's datasets, groups, organizations or other objects such as tags. Only free tags (tags that don't belong to a vocabulary) are returned. License list then returns the licenses available for datasets available on the portal. The second group then focuses on the searching process, i.e., search for packages or sources matching a query satisfying a given search criteria. This action accepts Apache Solr search query parameters and returns a dictionary of results, including the related datasets that match the search criteria, a search count and also facet information. Get an activity stream of all recently added or changed packages on a portal. Get a list of the site's user accounts and the roles for members of groups and organizations.

Figure 1 shows groups of countries based on the size of their open data portal. Most countries offer between 100 and 500 datasets. When compared to the others, Canada, the United States and the UK opened up more of their datasets to the public. These results are in agreement with the rankings of the GODI and the ODB index. In the GODI's rank order, the highest level of openness exists in the UK, Denmark and France. In the ODB's 2015 rank order, the highest level of openness exists in the UK, in the United States and Sweden.

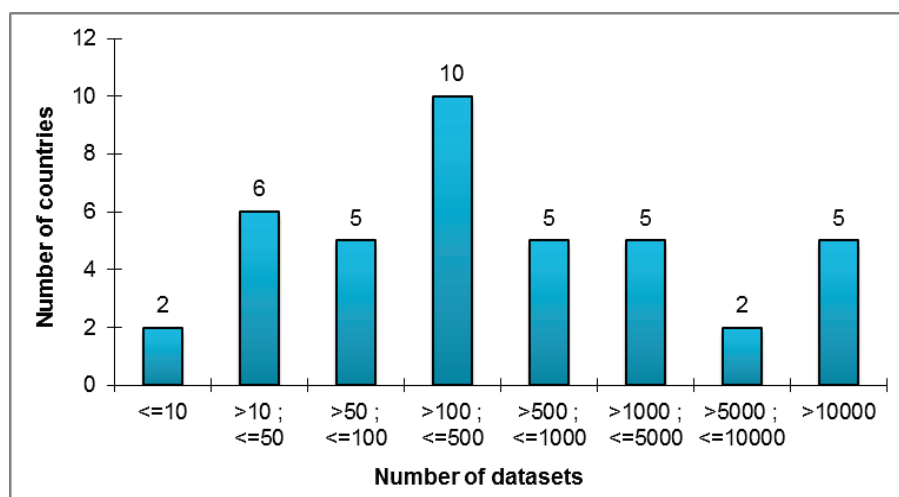


Figure 1: Histogram of numbers of available datasets

In the Figure 2, there are numbers of thematic groups available on the portals, which can be used to divide datasets into categories. However, this function is not compulsory, thus, 20 open data portals do not offer this function. More than 20 groups are offered by Australia, Brazil, Paraguay, Romania and Sweden. Figure 3 then shows the number of tags that are associated with related datasets. Figure 4 shows groups of countries based on the number of organizations, which participate in the open data portal. Generally, up to 50 organizations offered their data on these portals. Datasets from more than 250 organizations are available in the open data portal of Finland, Sweden and the UK. In both these figures, 10 open data portals do not offer this function through the API. Numbers of licenses for

open data publication and re-use are depicted in the Figure 5. Total of 15 portals do not offer this function through the API and 24 portals use only 15 default licenses (as D in the Figure 5), which are offered by CKAN.

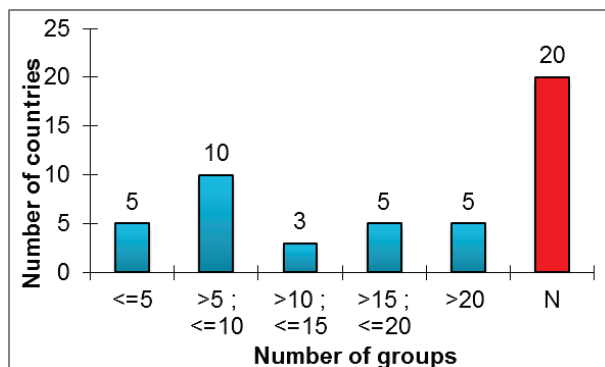


Figure 2: Histogram of numbers of thematic categories

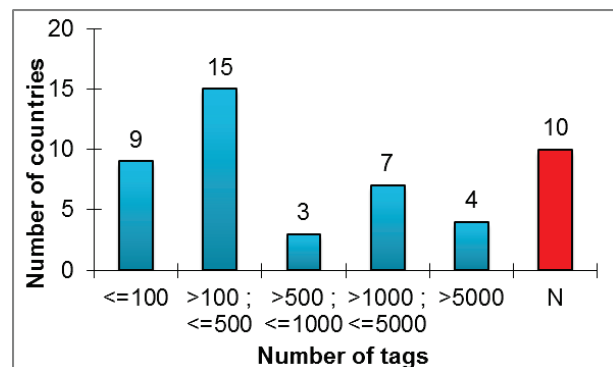


Figure 3: Histogram of numbers of tags

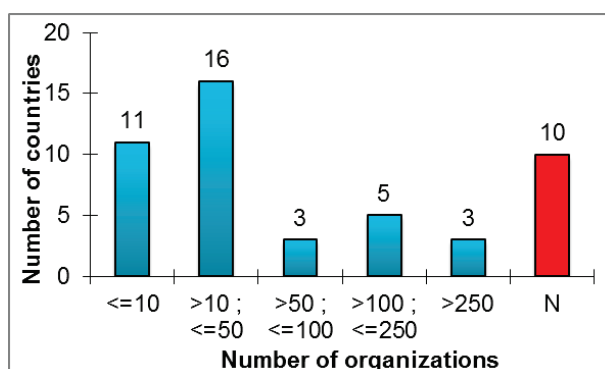


Figure 4: Histogram of numbers of organizations

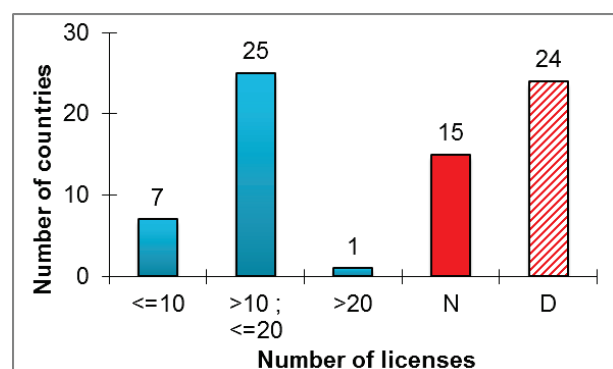


Figure 5: Histogram of numbers of licenses

The most widely used licenses are: Open Data Commons Public Domain Dedication and License (PDDL), Open Data Commons Open Database License (ODbL), Creative Commons CCZero and UK Open Government License (OGL). They are all available in 27 portals. Open Data Commons Attribution License, GNU Free Documentation License, Creative Commons Attribution and Creative Commons Non-Commercial then all offer 26 portals. Creative Commons Attribution Share-Alike is available in 25 portals. Only eight countries created their own licenses focusing on their legal environment. They are: Australia (three own licenses), Austria (three licenses), Canada (four licenses), Estonia (two licenses), Finland (nine licenses), Island (two licenses), Netherlands (eight licenses), Poland (six licenses), Romania and Uruguay, both with one own license. However, only selected datasets are published with one of these licenses. This issue was already mentioned in [42].

Also 39 countries offer the package search function through their API, 38 countries offer the resource search function and 38 countries offer the function to get an activity stream of all recently added or changed packages on the portal. Furthermore, the function user list should be available only to an authenticated user. However, as can be seen from the Figure 6, only 17 countries require the authentication of the user. The other countries make available information about their users such as name, created, email hash, activity streams email notifications, state, number of edits, number administered packages, etc. Finally, there are three basic member roles on each portal, which is powered by CKAN. These are: admin, editor and member.

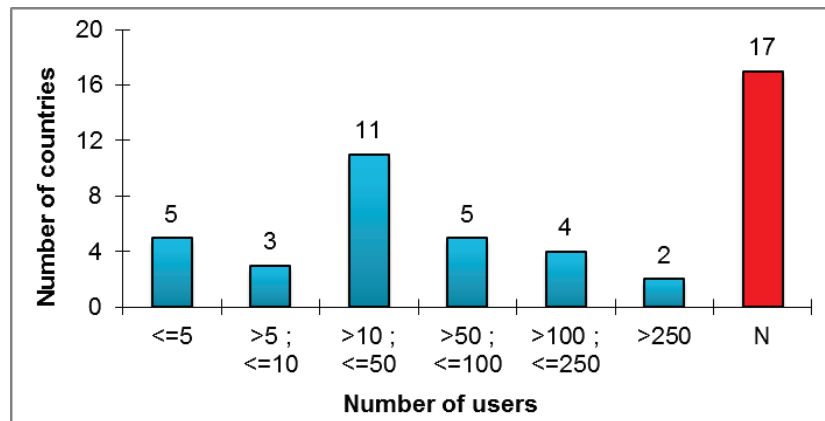


Figure 6: Histogram of users registered in the portal

## 4.2 Evaluating the Quality of Open Data Portals Using Benchmarking Framework

Based on the literature review conducted in the previous parts, three technological changes had a direct impact on the proposal of the benchmarking framework. The first technological change is the growth of broadband technologies and the speed of computational devices. They allow public sector institutions to interact with larger databases and develop services to personalize search and use their data more efficiently. The second technological change is the widespread use of social media and wiki platforms to create content, exchange ideas and best practices, or share, publish, discuss and collaborate with information. The use of these technologies by public sector institutions, citizens and businesses transforms the use of government data and the relationship between the stakeholders. The last technological change is the emergence of big data technologies to analyze, collect, and process large amounts of government data, which led to increase the potential uses of government data and their publishing and sharing. This change positively affected mashups of these data and improved collaboration using new and more reliable data.

The benchmarking framework presented in the Table 1 is based on the systematic literature review and authors' experiences and knowledge gained in the first part of the open data portals' evaluation. It follows the perspective of quality dimensions and metrics defined by Batini et al. [2]. The proposed framework is divided into two parts. The first one focuses on the general characteristics that consist of technical dimension, availability and access dimension, and communication and participation dimension. The second one evaluates the general characteristics of datasets and their metadata quality. In total, 28 complete criteria (metrics) are defined. This framework was firstly introduced in [30], but based on the discussion and feedback received, it was further modified. Some metrics were defined and described more detailed and one criterion for the general characteristics of datasets was added.

Each criterion was converted to a question to be included in a questionnaire to be distributed to users, e.g.: *This open data portal provides information about the authority, which hosts the portal and the governance model or institutional framework supporting data provision models.* These 28 questions are evaluated on a five point Likert scale to measure agreement or disagreement with such a statement (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). Each question can provide a score from one to five points, for a total score ranging from 28 to 140.

The evaluation questionnaire was initially tested by two associate professors highly experienced in quantitative research in the information system domain. They found it understandable and did not report any important problems. Then, 10 postgraduate students were trained in the capabilities of open data portals and their quality aspects. The training session, which lasted about one hour, consisted of two blocks: theoretical and practical. The theoretical block included the PowerPoint presentation about open data portals, evaluation requirements and benchmark criteria to be used for the evaluation. The presentation was followed by a question and answer session. The practical block included a sample procedure showing how to perform the evaluation. After the end of this training, they all filled the questionnaires in paper form. The users' evaluation data collected through the questionnaires were then processed to obtain the mean value for each portal. The results are then presented in the Table 2. In the first three columns, there are all evaluated dimensions for the general characteristics of the portal, followed by the sum of these scores. The average score for the general characteristics of the dataset can be found in the fifth column, followed by the overall score.

Table 1: The benchmarking framework for open data portals quality evaluation

<b>I. General characteristics of open data portal</b>	
List of metrics	Description of the requirements for the quality evaluation
<b>1. Technical dimension</b>	
1.1 Authority and responsibility	Portals should provide information about the authority, which hosts the portal and the governance model or institutional framework supporting data provision models [32], [44], [47].
1.2 Data management system	Portals should provide information about the data management system, which is used to power the portal [25], [33], [44].
1.3 Language	Portals should offer more language versions to gain more users (attention) and improve the overall quality of this portal [8], [42].
1.4 Free of charge	Portals should provide that all datasets and services are available free of charge and without any restrictions under open licenses [40], [52].
<b>2. Availability and access dimension</b>	
2.1 Number of datasets	Portals should provide the number of datasets they include [25], [33].
2.2 Number of applications (re-uses)	Portals should provide number of applications developed based on the open data re-used [47].
2.3 Search engine (filter)	Portals should adopt and make visible an overall organization structure and provide strong dataset search capabilities and selection tools using different criteria for browsing through categories and browsing through filters [8], [27], [32], [33], [42], [47].
2.4 API	Portals should provide API for stakeholders to develop applications using open data [8], [33], [42], [47].
2.5 User account	Portals should support user account creation in order to personalize views and information shown [8].
2.6 Thematic categories	Portals should provide thematic categories of the datasets provided by the portal [27], [42]. The portal should clearly distinguish categories (themes) from tags (keywords) [33]. Same tags should be used to classify data of the same type and category [8], [27], [32].
2.7 Tags (keywords)	
<b>3. Communication and participation dimension</b>	
3.1 Forum (feedback)	Portals should provide an opportunity to submit feedback on the data from the users to providers and forum to discuss and exchange ideas among the users [8], [42], [47], [57].
3.2 Request form	Portals should provide a form to request or suggest new type or format type of open data [27], [33], [57].
3.3 Help (usability)	Portals should include high quality of documentation and help functionality to learn how to use the portal and improve the usability [8], [44], [47].
3.4 Frequently Asked Questions (FAQ)	Portals should provide a FAQ section to help resolve any potential issues [47].
3.5 Social media	Portals should be connected to a social media platform to create a social distribution channel for open data. OGD users and providers can inform each other about what they did with and learned from a dataset [42], [47], [57].
<b>II. General characteristics of dataset</b>	
1. Title and description	Datasets should be provided together with their description and also how and for what purpose they were collected [8], [27], [40].
2. Publisher	Datasets should be provided together with their publisher to verify authenticity of their source [33].
3. Release date and up to date	Datasets should be explicitly associated with a specific time or period tag. All information in the dataset should be up to date [27], [32], [33], [40], [52].
4. License	Datasets should provide license information related to the use of the published datasets. Datasets that doesn't explicitly have an open license are not open data [33], [40], [42], [52].
5. Geographic coverage	Datasets should be determined if the coverage of data is on the national, regional or local level [33], [42].
6. Dataset URL	Datasets URL should be available for each dataset [33], [42].
7. Dataset (file) size	Datasets (file) size should be available [25], [33].
8. Number of views (visits)	Total number of online views should be available for a dataset [47].
9. Number of downloads	Total number of downloads should be available for a dataset [47].
10. Machine-readable formats	Datasets should be provided in formats that are as convenient, easy to analyze and modifiable as downloadable files in well-known formats [8], [27], [33], [40], [47], [52].
11. Visualizations	Datasets visualization capabilities should be provided, e.g., as visualizations in charts or visualizations in maps [42].
12. User rating and discussion message	Datasets should provide capabilities allowing to collect user ratings and comments on a dataset or to discuss conclusions based on data use [8], [57].

Table 2: The results of the open data portals quality evaluation

Country	Open data portal	I.				II.	Overall score
		1.	2.	3.	Sum		
UK	Site 4	19	34	23	76	53.6	<b>129.6</b>
India	Site 5	19.5	29.8	21.4	70.7	55.3	<b>126</b>
United States	Site 6	18.5	34	20.1	72.6	52	<b>124.6</b>
Australia	Site 7	18.2	33.1	19.8	71.1	49.3	<b>120.4</b>
Austria	Site 8	16.7	31.5	20.6	68.8	45.4	<b>114.2</b>
Canada	Site 9	17.9	29.4	21	68.3	43	<b>111.3</b>
France	Site 10	16.5	29.9	15.3	61.7	49	<b>110.7</b>
Russia	Site 11	12.4	28.2	21.3	61.9	48.7	<b>110.6</b>
Moldova	Site 12	17.6	29.4	15.5	62.5	46.9	<b>109.4</b>
Croatia	Site 13	15.5	30.8	20.6	66.9	40.3	<b>107.2</b>
Germany	Site 14	15.3	33	17.2	65.5	41.1	<b>106.6</b>
Paraguay	Site 15	17.5	32.2	19.1	68.8	37	<b>105.8</b>
Greece	Site 16	15	29.3	16.2	60.5	43.7	<b>104.2</b>
Finland	Site 17	15.7	26	11.1	52.8	51.4	<b>104.2</b>
Chile	Site 18	16.1	27.5	11.5	55.1	48.8	<b>103.9</b>
Brazil	Site 19	15.3	28.4	17.2	60.9	41.6	<b>102.5</b>
Switzerland	Site 20	18.5	29.2	15.3	63	39.4	<b>102.4</b>
New Zealand	Site 21	15.1	28.9	18.6	62.6	39.7	<b>102.3</b>
Ireland	Site 22	18.1	28.2	15.9	62.2	39.9	<b>102.1</b>
Nepal	Site 23	15.3	24.4	17	56.7	43.5	<b>100.2</b>
Taiwan	Site 24	12	25.5	19.4	56.9	43	<b>99.9</b>
Portugal	Site 25	14.6	26.4	17.4	58.4	41.2	<b>99.6</b>
Tanzania	Site 26	15.4	27.4	17.8	60.6	38.2	<b>98.8</b>
Japan	Site 27	16.3	26.6	10.4	53.3	45.4	<b>98.7</b>
Belarus	Site 28	11.6	31.2	15.4	58.2	40.2	<b>98.4</b>
Bulgaria	Site 29	18.7	23.8	16.8	59.3	38.6	<b>97.9</b>
South Korea	Site 30	15.2	26.9	17.2	59.3	38.5	<b>97.8</b>
Uruguay	Site 31	15.8	32.8	10.4	59	38.7	<b>97.7</b>
Sweden	Site 32	18.3	26.5	12.9	57.7	39.9	<b>97.6</b>
Spain	Site 33	12.8	28.2	17.2	58.2	39.1	<b>97.3</b>
Italy	Site 34	15.1	22.7	19.1	56.9	39.6	<b>96.5</b>
Colombia	Site 35	11.4	23.7	18.9	54	42.4	<b>96.4</b>
Indonesia	Site 36	14.3	27.6	15.4	57.3	38.4	<b>95.7</b>
Singapore	Site 37	13.1	29.2	11.7	54	40.6	<b>94.6</b>
Slovakia	Site 38	18.5	26	12	56.5	37.9	<b>94.4</b>
El Salvador	Site 39	13.4	23.1	20.5	57	37.1	<b>94.1</b>
Romania	Site 40	18.2	26.1	10.2	54.5	39.2	<b>93.7</b>
Morocco	Site 41	17.9	30	10.9	58.8	33.9	<b>92.7</b>
Netherlands	Site 42	13.3	27.5	10.5	51.3	41.4	<b>92.7</b>
Philippines	Site 43	15.4	28.9	13	57.3	35.2	<b>92.5</b>
Ghana	Site 44	15.3	22.5	18.4	56.2	36.2	<b>92.4</b>
Denmark	Site 45	14.3	24.9	18	57.2	35	<b>92.2</b>
Pakistan	Site 46	17.1	28.6	10.6	56.3	34.9	<b>91.2</b>
Norway	Site 47	13.2	32	12.4	57.6	33.1	<b>90.7</b>
Poland	Site 48	16.1	26.3	10.5	52.9	37.7	<b>90.6</b>
Kenya	Site 49	16.8	23.9	20.7	61.4	28.8	<b>90.2</b>
Sri Lanka	Site 50	15.2	22.6	19.3	57.1	32.9	<b>90</b>
Macedonia	Site 51	13.2	23.1	13.5	49.8	38.5	<b>88.3</b>
Belgium	Site 52	14.5	19.9	15.7	50.1	38.1	<b>88.2</b>
Thailand	Site 53	10.7	21.5	19.5	51.7	36.5	<b>88.2</b>
Czech Republic	Site 54	16.4	26.5	6.4	49.3	38.1	<b>87.4</b>
Uzbekistan	Site 55	15.7	25.4	12.8	53.9	33.2	<b>87.1</b>
Kazakhstan	Site 56	17.6	25.1	10.5	53.2	33.5	<b>86.7</b>
Cyprus	Site 57	14.5	18.7	17.4	50.6	35.4	<b>86</b>
Serbia	Site 58	16.6	25	6.4	48	37.5	<b>85.5</b>
Ukraine	Site 59	13.3	25.7	14.9	53.9	31.2	<b>85.1</b>
Costa Rica	Site 60	13.7	23.1	13.4	50.2	34.7	<b>84.9</b>
Uganda	Site 61	14.5	22.4	12.6	49.5	34.7	<b>84.2</b>
Saudi Arabia	Site 62	15.9	19.2	15.1	50.2	33.8	<b>84</b>

*Table 2: continuation*

Hong Kong	Site 63	16.8	18.2	15.1	50.1	28.9	<b>79</b>
Mexico	Site 64	11.7	22.4	13.3	47.4	30.2	<b>77.6</b>
Brunei	Site 65	12.6	21.8	15.4	49.8	26.6	<b>76.4</b>
Malaysia	Site 66	13	19.7	10.6	43.3	30.5	<b>73.8</b>
Lithuania	Site 67	8.1	26.2	10.3	44.6	27.5	<b>72.1</b>
Burkina Faso	Site 68	13.5	21.5	10.4	45.4	26.5	<b>71.9</b>
Tunisia	Site 69	12.4	15	19	46.4	23.3	<b>69.7</b>
Israel	Site 70	10.1	15.1	13.4	38.6	30.1	<b>68.7</b>

## 5 Findings and Discussion

The UK received the highest score among the evaluated countries, which is in the line with the GODI and the ODB index. These indices and related reports rank the UK as the world's most developed open data ecosystem. It is also in agreement with the findings of Heimstädt et al. presented in [18]. The UK is followed by India, which provides the highest quality datasets, the United States and Australia. Most of the evaluated portals are powered by CKAN, however, not always by the latest version. Therefore, they received a low score such as in the case of the Czech Republic. However, on the other hand, the best results were achieved by countries, where the open data portal is powered by CKAN or DKAN, except of the largest advanced economies such as France, Germany or the United States.

Basically, on the best performing open data portals, i.e., the UK, India, the United States, Australia and Austria, there were found various functions that can be conducted, such as searching, requesting, viewing, downloading, analyzing, combining, visualizing and discussing different types of datasets as well as open data trends and concepts. For instance, an open data user may change the language in the main menu, search for data, then filter the results and then zoom in on the dataset and then analyze it in detail. Most of these portals also offer a personalized approach to use the available functions by creating a user account. The evaluation also showed that some portals provide advanced engagement capability and features such as: blogs to facilitate the sharing of sources and knowledge, a forum to facilitate discussion on related applications and technology, and community groups. Overall, some of these open data portals constitute a single point-of-access for a broader open government policy and initiatives, opening up government data and new technologies, such as the current government use of social media platforms. Furthermore, although all portals provide some kind of metadata for each dataset, their usefulness is sometimes limited by the fact that there is no complete listing of possible metadata fields and their description is provided, several metadata fields remain empty or it is simply difficult to find them on the portal.

From the evaluation presented in the previous sections, it may be concluded that most of the open data portals in the world have achieved some significant steps towards opening a number of interesting and useful datasets. Therefore, the findings presented in this paper are in the line with Alexopoulos et al. [1], which reported that the existing first generation of data portals offers mainly basic functionalities for searching and downloading data by the users of these data, and for uploading data by their providers. The majority of these portals offer simple free-text search and theme-browsing functions for the discovery of datasets. Only some of the best performing open data portals have recently taken advantage of Semantic Web by providing semantically enriched discovery services and only a few of them provide functionality to view datasets on a map, include dataset's rating and commenting or various types of charts. However, there are no functionalities for processing the datasets in order to improve them, adapt them to specialized needs, or link them to other datasets (public or private), and then for uploading-publishing new versions of them, or for uploading users' own datasets.

Thus, this paper indicates that there are some important improvements that should be made, in order to enhance openness in some countries, and increase the social and economic value that can be generated from them. Some recommendations in this direction are provided as follow.

As the findings of the quality evaluation showed, the quality of open data portals is affected by the version of the data management system, especially in the case of CKAN, and open data portals powered by CKAN or DKAN achieved better scores than the other portals. CKAN is released under several versions, which differ from each other in terms of features and service level. Each version offers various functionalities, which may improve the quality of open data portals as well as related datasets and their metadata. Therefore, it may be suggested that the other portals should migrate their services to CKAN or DKAN, especially in the case of developing countries.

Also, more datasets should be opened, from a wider range of thematic categories, such as more economic/financial datasets concerning government spending, which should lead to higher government transparency and accountability, and on economic activity of businesses. At the same time, more emphasis should be placed on opening datasets of some important categories that have been neglected, such as employment related datasets, agriculture and tourism related datasets or environment and planning related datasets, which are important for citizens' quality of life. More emphasis should be placed on the use of structured and machine-readable file formats in publishing datasets, and metadata (adopting existing metadata standards), and on the support of RDF and also Protocol and RDF Query Language (SPARQL), which will enable more effective browsing and discovery of datasets, and also linking and

combining open data from multiple sources, leading to a big increase of their usefulness and value for various stakeholders. Furthermore, there need to be interactions with data providers, as users can request data from data providers and provide them with feedback after they have used these data. Information about the licenses that are connected to the use of certain datasets is also important in open data ecosystems, as open data users need to know whether the license allows them to use the data in the way that they want to use them. Also, the information about registered users should be available only to the authenticated user. This function may not be available through the API. Finally, interoperability of evaluated portals can be improved by providing metadata about shared identifiers and vocabularies and by reusing related elements. This lack of harmonization increases the need for quality standards.

Authorities responsible for these open data portals must therefore consider the specific needs of those looking for accountability-related data, and provide structures and mechanisms to address them [32]. At the same time the functionality of the existing OGD sources should be enhanced, providing more advanced tools mainly for data discovery (so that potential users can find more easily and quickly the datasets they are interested in), data visualization (for instance on maps and charts, so that potential users can easily and quickly get a first understanding about the dataset, and decide whether it is worth continuing with a more detailed analysis) and users' feedback (so that OGD users can provide feedback to their providers, about the quality of the datasets they have used, existing weaknesses and necessary improvements, and needs for additional datasets - as the collaboration between OGD users and providers has been recognized as critical for the generation of value from them, e.g. in [21], [22], [32] or [56], [58]).

In contrast to [30], [31], [33], [42], [44], [48], [51], this paper distinguishes between the quality of a single portal and the quality of datasets and their metadata on this portal. Furthermore, it uses bigger research sample than these studies. Finally, it presents important findings about improving the quality of these portals, especially in the area of security and information protection.

Limitations and reliability of the results presented in this paper may be affected by the related initiatives, legislation and applicable regulations, which do not allow the use of data for purposes other than those regulated and for which the data were collected. More precisely, only data complying with these requirements may be found on the evaluated open data portals. However, the framework was proposed with this in mind as it evaluates mostly the metadata description of related datasets. Also, the benchmarking framework does not evaluate any functions, which require the authenticated user. Furthermore, the evaluated portals have complex (and not always clear) organizational and categorization structures that are continuously evolving, which hinders any analysis conducted from an ordinary user point-of-view (that is, without inside knowledge of each portal). All the criteria in the benchmarking framework are considered of equal relevance. However, some of them may be more important than others. Although there is no comprehensive prioritization of these requirements in the literature, it may affect the final ranking. For example, if the general characteristics of open data portal would have the weight of 2/3 and the general characteristics of dataset would have the weight of 1/3, then the final ranking would be: UK, United States, India, Australia, Austria, Canada, Paraguay, Croatia, Russia, and France.

Another very important point is the question of users used to evaluate the selected portals. In this case, only the postgraduate students participated in the evaluation. However, there are also other stakeholders such as the local and state governments, businesses, citizens, etc. Therefore, the future research should be focused on the further study of these stakeholders and their requirements and what they think is the value of these open data portals. Also, the selected open data portals on the regional or local level of the public sector may be evaluated. Finally, the need for open data quality standards is once again being recognized.

## 6 Conclusions

This paper presents the benchmarking framework for the quality evaluation of open data portals on the national level as well as in-depth review of the issues, challenges and opportunities associated with these portals. For this purpose, the systematic literature review and methods of content and multi-dimension analysis together with quantitative techniques were used.

Open data portals are the interfaces between government data on one side and re-users on the other side. Basically, each portal should have a clean look with a search bar on the homepage, information about the authority, which hosts the portal, and the content should be written simply and structured into categories and also tags. Apart from making data available to stakeholders, the portal should also aim to engage citizens' ideas and feedback. The efficient development of these portals makes it necessary to evaluate their quality systematic, in order to understand them better and assess the various types of value they generate. Therefore, the proposed framework is the main theoretical contribution of this paper. However, the quality evaluation of open data portals is not only important for understanding the value the portals generate, but also for further improvements. Findings from evaluations can be used to further improve open datasets and open data portals, so that they become more useful and so that adoption will increase. Finally, there are also mentioned several observations and recommendations in terms of practical contributions of open data portals development drawn from the presented evaluation.

However, the number of datasets online and the sophistication of open data portals and their functions differ, reflecting the lack of harmonization and the need for quality standards. In particular, the UK, India and the United States have published many datasets and launched advanced websites. It is also clear that the quality of data and the language settings are very important keys, but the main task is to provide with legal manner, quality standards and up-dated data for those have the right to use it as tool, info, metadata etc. Therefore, the results of the quality evaluation using new benchmarking framework can be used to improve the quality of open data portals as well as data infrastructures, strategies and initiatives.

## Acknowledgments

The authors would like to thank the reviewers for taking the time to assess this paper and for providing valuable critique which helped to improve this paper. The authors would like to acknowledge the assistance of the Student Grant Competition of the University of Pardubice No. SGS\_2016\_023.

## Websites List

Site 1: Data portals: A comprehensive list of open data portals from around the world  
<http://dataportals.org/>

Site 2: Catalog of open data portals  
<http://opengeocode.org/opendata/>

Site 3: CKAN instances around the world  
<http://ckan.org/instances/>

Site 4: National open data portal of the UK  
<https://data.gov.uk/>

Site 5: National open data portal of India  
<https://data.gov.in/>

Site 6: National open data portal of United States  
<https://www.data.gov/>

Site 7: National open data portal of Australia  
<http://data.gov.au/>

Site 8: National open data portal of Austria  
<https://www.data.gv.at/>

Site 9: National open data portal of Canada  
<http://open.canada.ca/>

Site 10: National open data portal of France  
<https://www.data.gouv.fr/fr/>

Site 11: National open data portal of Russia  
<http://data.gov.ru/>

Site 12: National open data portal of Moldova  
<http://data.gov.md/>

Site 13: National open data portal of Croatia  
<http://data.gov.hr/>

Site 14: National open data portal of Germany  
<https://www.govdata.de/>

Site 15: National open data portal of Paraguay  
<http://datos.org.py/>

Site 16: National open data portal of Greece  
<http://data.gov.gr/>



Site 17: National open data portal of Finland

<https://www.opendata.fi/>

Site 18: National open data portal of Chile

<http://datos.gob.cl/>

Site 19: National open data portal of Brazil

<http://dados.gov.br/>

Site 20: National open data portal of Switzerland

<http://opendata.admin.ch/>

Site 21: National open data portal of New Zealand

<https://data.govt.nz/>

Site 22: National open data portal of Ireland

<http://data.gov.ie/>

Site 23: National open data portal of Nepal

<http://data.opennepal.net/>

Site 24: National open data portal of Taiwan

<http://data.gov.tw>

Site 25: National open data portal of Portugal

<http://www.dados.gov.pt/>

Site 26: National open data portal of Tanzania

<http://opendata.go.tz/>

Site 27: National open data portal of Japan

<http://www.data.go.jp/>

Site 28: National open data portal of Belarus

<https://opendata.by/>

Site 29: National open data portal of Bulgaria

<http://opendata.government.bg/>

Site 30: National open data portal of South Korea

<https://www.data.go.kr/>

Site 31: National open data portal of Uruguay

<https://catalogodatos.gub.uy/>

Site 32: National open data portal of Sweden

<http://opnadata.se/>

Site 33: National open data portal of Spain

<http://datos.gob.es/>

Site 34: National open data portal of Italy

<http://www.dati.gov.it/>

Site 35: National open data portal of Colombia

<https://www.datos.gov.co/>

Site 36: National open data portal of Indonesia

<http://data.go.id/>

Site 37: National open data portal of Singapore

<http://data.gov.sg/>

Site 38: National open data portal of Slovakia

<https://data.gov.sk/>

Site 39: National open data portal of El Salvador  
<http://www.datoselsalvador.org/>

Site 40: National open data portal of Romania  
<http://data.gov.ro/>

Site 41: National open data portal of Morocco  
<http://data.gov.ma/>

Site 42: National open data portal of the Netherlands  
<https://data.overheid.nl/>

Site 43: National open data portal of Philippines  
<http://data.gov.ph/>

Site 44: National open data portal of Ghana  
<http://data.gov.gh/>

Site 45: National open data portal of Denmark  
<http://data.digitaliser.dk/>

Site 46: National open data portal of Pakistan  
<http://data.org.pk/>

Site 47: National open data portal of Norway  
<http://data.norge.no/>

Site 48: National open data portal of Poland  
<https://danepubliczne.gov.pl/>

Site 49: National open data portal of Kenya  
<https://www.opendata.go.ke/>

Site 50: National open data portal of Sri Lanka  
<https://www.data.gov.lk/>

Site 51: National open data portal of Macedonia  
<http://www.otvorenipodatoci.gov.mk/>

Site 52: National open data portal of Belgium  
<http://data.gov.be>

Site 53: National open data portal of Thailand  
<http://data.go.th/>

Site 54: National open data portal of the Czech Republic  
<https://datahub.io/organization/cz-ckan-net>

Site 55: National open data portal of Uzbekistan  
<https://data.gov.uz/uz>

Site 56: National open data portal of Kazakhstan  
<https://data.egov.kz/>

Site 57: National open data portal of Cyprus  
<http://www.data.gov.cy/>

Site 58: National open data portal of Serbia  
<http://rs.ckan.net/>

Site 59: National open data portal of Ukraine  
<http://data.gov.ua/>

Site 60: National open data portal of Costa Rica  
<http://datosabiertos.presidencia.go.cr/home>

Site 61: National open data portal of Uganda

<http://www.data.ug/>

Site 62: National open data portal of Saudi Arabia

<http://www.data.gov.sa/>

Site 63: National open data portal of Hong Kong

<https://data.gov.hk/>

Site 64: National open data portal of Mexico

<http://datos.gob.mx/>

Site 65: National open data portal of Brunei

<http://www.data.gov.bn/>

Site 66: National open data portal of Malaysia

<http://data.gov.my/>

Site 67: National open data portal of Lithuania

<http://opendata.gov.lt/>

Site 68: National open data portal of Burkina Faso

<http://alpha.data.gov.bf/>

Site 69: National open data portal of Tunisia

<http://data.gov.tn/>

Site 70: National open data portal of Israel

<https://data.gov.il/>

## References

- [1] C. Alexopoulos, et al., Designing a second generation of open data platforms: Integrating open data and social media, in Proceedings of the 13th IFIP WG 8.5 International Conference (EGOV 2014), 2014, pp. 230-241.
- [2] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, Methodologies for data quality assessment and improvement, ACM Computing Surveys, vol. 41, no. 3, pp. 1-52, 2009.
- [3] R. Bílková, R. Máchová and M. Lněnička, M., Evaluating the impact of open data using partial least squares structural equation modeling, Scientific Papers of the University of Pardubice - Series D, Faculty of Economics and Administration, vol. 22, no. 34, pp. 29-41, 2015.
- [4] K. Braunschweig, J. Eberius, M. Thiele, and W. Lehner, The state of open data - limits of current open data platforms, in Proceedings of the 21st World Wide Web Conference 2012, Web Science Track at WWW'12, ACM, New York, 2012, pp. 1-6.
- [5] S. Buchholtz, M. Bukowski and A. Śniegocki, Big and Open Data in Europe: A Growth Engine or a Missed Opportunity. Varšava: demosEUROPA, 2014.
- [6] I. Caballero, M. Serrano and M. Piattini, A data quality in use model for big data, in Advances in Conceptual Modeling (M. Indulska and S. Purao, Eds.). Atlanta, GA: Springer, pp. 65-74, 2014.
- [7] C. Calero, A. Caro and M. Piattini, An applicable data quality model for web portal data consumers, world wide web, vol. 11, no. 4, pp. 465-484, 2008.
- [8] Y. Charalabidis, E. Loukis and C. Alexopoulos, Evaluating second generation open government data infrastructures using value models, in Proceedings of the 47th Hawaii International Conference on System Sciences, IEEE, Hilton Waikoloa, 2014, pp. 2114-2126.
- [9] M. Chen, S. Mao and Y. Liu, Big data: A survey, Mobile Networks and Applications, vol. 19, no. 2, pp. 171-209, 2014.
- [10] P. Colpaert, S. Joye, P. Mechant, E. Mannens, and R. van de Walle, The 5 stars of open data portals, in Proceedings of the 7th International Conference on Methodologies, Technologies and Tools Enabling E-Government (MeTTeG13), University of Vigo, Spain, 2013, pp. 61-67.
- [11] D. Cowan, P. Alencar and F. McGarry, Perspectives on open data: Issues and opportunities, in Proceedings of the 2014 IEEE International Conference on Software Science, Technology and Engineering, IEEE, Ramat Gal, Israel, 2014, pp. 24-33.
- [12] R. Cyganiak and F. Maali. (2015, May) Use cases and requirements for the data catalog vocabulary. W3C. [Online]. Available: <https://dvcs.w3.org/hg/gld/raw-file/default/dcat-ucr/index.html>
- [13] S. S. Dawes and N. Helbig, Information strategies for open government: Challenges and prospects for deriving public value from government transparency, in Proceedings of the 9th IFIP WG 8.5 International Conference (EGOV 2010), Springer, Heidelberg, 2010, pp. 50-60.
- [14] I. Ermilov, et al., Linked open data statistics: Collection and exploitation, in Knowledge Engineering and the Semantic Web (P. Klinov and D. Mourontsev, Eds.). Berlin Heidelberg: Springer, 2013, pp. 242-249.

- [15] C. P. Geiger and J. von Lucke, Open government and (linked) (open) (government) (data), *eJournal of eDemocracy & Open Government*, vol. 4, no. 2, pp. 265-278, 2012.
- [16] H. S. Hansen, L. Hvingel and L. Schrøder, Open government data - A key element in the digital society, in *Technology-Enabled Innovation for Democracy, Government and Governance* (A. Ko, C. Leitner, H. Leitold, and A. Prosser, Eds.). Berlin: Springer, 2013, pp. 167-180.
- [17] T. Heath and C. Bizer, Linked data: Evolving the web into a global data space, *Synthesis Lectures on the Semantic Web*, vol. 1, no. 1, pp. 1-136, 2011.
- [18] M. Heimstädt, F. Saunderson and T. Heath, Conceptualizing open data ecosystems: A timeline analysis of open data development in the UK, in *CeDEM14: Proceedings of the International Conference for E-Democracy and Open Government 2014*, Edition Donau-Universität Krems, Krems, 2014, pp. 245-255.
- [19] B. Hyland and D. Wood, The joy of data-a cookbook for publishing linked government data on the web, in *Linking Government Data* (D. Wood, Ed.). New York: Springer, 2011, pp. 3-26.
- [20] M. Janssen, Y. Charalabidis and A. Zuiderwijk, Benefits, adoption barriers and myths of open data and open government, *Information Systems Management*, vol. 29, no. 4, pp. 258-268, 2012.
- [21] T. Jetzek, M. Avital and N. Bjørn-Andersen, Generating value from open government data, in *Proceedings of 34th International Conference on Information Systems: ICIS 2013*, Bepress, Berkeley, 2013, pp. 1-20.
- [22] T. Jetzek, M. Avital and N. Bjørn-Andersen, Generating sustainable value from open data in a sharing society, in *Proceedings of IFIP WG 8.6 International Conference on Transfer and Diffusion of IT (TDIT 2014)*, Berlin, Springer, 2014, pp. 62-82.
- [23] E. Kalampokis, E. Tambouris and K. Tarabanis, Linked open government data analytics, in *Proceedings of the 12th IFIP WG 8.5 International Conference (EGOV 2013)*, Springer, Heidelberg, 2013, pp. 99-110.
- [24] E. Kalampokis, E. Tambouris and K. Tarabanis, Open government data: A stage model, in *Proceedings of the 10th IFIP WG 8.5 International Conference (EGOV 2011)*, Springer, Heidelberg, 2011, pp. 235-246.
- [25] M. Kostovski, M. Jovanovik and D. Trajanov, Open data portal based on semantic web technologies, in *Proceedings of the 7th South East European Doctoral Student Conference*, University of Sheffield, Greece, 2012, pp. 1-13.
- [26] J. Kucera and D. Chlapek, Benefits and risks of open government data, *Journal of Systems Integration*, vol. 5, no. 1, pp. 30-41, 2014.
- [27] J. Kučera, D. Chlapek and M. Nečaský, Open government data catalogs: Current approaches and quality perspective, in *Technology-Enabled Innovation for Democracy, Government and Governance* (A. Ko, C. Leitner, H. Leitold, and A. Prosser, Eds.). Berlin: Springer, 2013, pp. 152-166.
- [28] T. Lebo, et al., Producing and using linked open government data in the TWC LOGD portal, in *Linking Government Data* (D. Wood, Ed.). New York: Springer, 2011, pp. 51-72.
- [29] S. M. Lee, T. Hwang and D. Choi, Open innovation in the public sector of leading countries, *Management Decision*, vol. 50, no. 1, pp. 147-162, 2012.
- [30] M. Lněnička, An in-depth analysis of open data portals as an emerging public e-service, *International Journal of Social, Behavioral, Educational, Economic and Management Engineering*, vol. 9, no. 2, pp. 589-599, 2015.
- [31] M. Lněnička and R. Máchová, Open (big) data and the importance of data catalogs and portals for the public sector, in *Proceedings in Global Virtual Conference: The 3rd International Global Virtual Conference (GV-CONF 2015)*, EDIS - Publishing Institution of the University of Zilina, Zilina, 2015, pp. 143-148.
- [32] R. P. Lourenço, An analysis of open government portals: A perspective of transparency for accountability, *Government Information Quarterly*, vol. 32, no. 3, pp. 323-332, 2015.
- [33] F. Maali, R. Cyganiak and V. Peristeras, Enabling interoperability of government data catalogues, in *Proceedings of the 9th IFIP WG 8.5 International Conference (EGOV 2010)*, Springer, Heidelberg, 2010, pp. 339-350.
- [34] R. Máchová and M. Lněnička, Exploring the emerging impacts of open data in the public sector, in *Proceedings of the 20th International Conference Current Trends in Public Sector Research*, Masaryk University, Brno, 2016, pp. 36-44.
- [35] R. Máchová and M. Lněnička, Reframing e-government development indices with respect to new trends in ICT, *Review of Economic Perspectives*, vol. 15, no. 4, pp. 383-411, 2015.
- [36] U. Maier-Rabler and S. Huber, Open: the changing relation between citizens, public administration, and political authority, *eJournal of eDemocracy & Open Government*, vol. 3, no. 2, pp. 182-191, 2011.
- [37] S. Martin, M. Foulonneau, S. Turki, and M. Ihadjadene, Open data: Barriers, risks and opportunities, in *Proceedings of the 13th European Conference on eGovernment (ECEG 2013)*, Academic Conferences and Publishing International Limited, Reading, 2013, pp. 301-309.
- [38] A. Marton, M. Avital and T. Blegind Jensen, Reframing open big data, in *Proceedings of the 21st European Conference on Information Systems*, Utrecht University, Netherlands, 2013, pp. 1-12.
- [39] H. Maylor and K. Blackmon, *Researching Business and Management*. New York: Palgrave-Macmillan, 2005.
- [40] Methodology. (2013, October) Global open data index by open knowledge. Global Open Data Index. [Online]. Available: <http://index.okfn.org/methodology/>
- [41] M. Petticrew and H. Roberts, *Systematic Reviews in the Social Sciences: A Practical Guide*, Malden: Blackwell Publishing, 2006.
- [42] M. Petychakis, O. Vasileiou, C. Georgjis, S. Mouzakis, and J. Psarras, A state-of-the-art analysis of the current public data landscape from a functional, semantic and technical perspective, *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 9, no. 2, pp. 34-47, 2014.

- [43] L. Reggi and C. A. Ricci, Information strategies for open government in Europe: EU regions opening up the data on structural funds, in Proceedings of the 10th IFIP WG 8.5 International Conference (EGOV 2011), Springer, Heidelberg, 2011, pp. 173-184.
- [44] D. S. Sayogo, T. A. Pardo and M. Cook, A framework for benchmarking open government data efforts, in Proceedings of the 47th Hawaii International Conference on System Sciences, IEEE, 2014, pp. 1896-1905.
- [45] M. Solar, G. Concha and L. Meijueiro, A model to assess open government data in public agencies, in Proceedings of the 11th IFIP WG 8.5 International Conference (EGOV 2012), Springer, Heidelberg, 2012, pp. 210-221.
- [46] J. M. Tien, Big data: Unleashing information, Journal of Systems Science and Systems Engineering, vol. 22, no. 2, pp. 127-151, 2013.
- [47] B. Ubaldi, Open government data: Towards empirical analysis of open government data initiatives, OECD, France, Working Papers on Public Governance, no. 22, OECD Publishing, 2013.
- [48] J. Umbrich, S. Neumaier and A. Polleres, Quality assessment & evolution of open data portals, in Proceedings IEEE International Conference on Open and Big Data, IEEE, Rome, 2015, pp. 1-8.
- [49] United Nations, United Nations E-government Survey 2014: E-Government for the Future We Want. New York: UN Publishing Section, 2014.
- [50] S. Van der Waal, et al., Lifting open data portals to the data web, in Linked Open Data - Creating Knowledge Out of Interlinked Data (S. Auer, V. Bryl and S. Tramp, Eds.). Cham: Springer International Publishing, 2014, pp. 175-195.
- [51] N. Verma and M. P. Gupta, Open government data: More than eighty formats, in Proceedings of the 9th International Conference on E-Governance (ICEG 2012), CSI, Cochin, India, 2012, pp. 207-216.
- [52] World Wide Web Foundation, Open Data Barometer Global Report: Second Edition. Washington DC: World Wide Web Foundation, 2015.
- [53] H.-C. Yang, C. S. Lin and P.-H. Yu, Toward automatic assessment of the categorization structure of open data portals, in Multidisciplinary Social Networks Research (L. Wang, S. Uesugi, I.-H. Ting, K. Okuhara, and K. Wang, Eds.). Berlin: Springer Heidelberg, 2015, pp. 372-380.
- [54] Z. Yang and A. Kankanhalli, Innovation in government services: The case of open data, in Grand Successes and Failures in IT: Public and Private Sectors (Y. K. Dwivedi, H. Z. Henriksen, D. Wastell, and R. De', Eds.). Berlin: Springer, 2013, pp. 644-651.
- [55] S. Zillner, et al. (2014) Public deliverable of the EU-project BIG (318062; ICT-2011.4.4). BIG. [Online]. Available: [http://big.atosresearch.eu/sites/default/files/content-files/deliverables/BIG\\_D2\\_3\\_2.pdf](http://big.atosresearch.eu/sites/default/files/content-files/deliverables/BIG_D2_3_2.pdf)
- [56] A. Zuiderwijk and M. Janssen, A coordination theory perspective to improve the use of open data in policy-making, in Proceedings of the 12th IFIP WG 8.5 International Conference (EGOV 2013), Springer, Heidelberg, 2013, pp. 38-49.
- [57] A. Zuiderwijk and M. Janssen, Participation and data quality in open data use: Open data infrastructures evaluated, in Proceedings of the 15th European Conference on E-Government 2015 (ECEG 2015), Academic Conferences and Publishing International Limited, Reading, UK, 2015, pp. 351-359.
- [58] A. Zuiderwijk, M. Janssen and C. Davis, Innovation with open data: Essential elements of open data ecosystems, Information Polity, vol. 19, no. 1, 2, pp. 17-33, 2014.