International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:9, No:2, 2015

An In-Depth Analysis of Open Data Portals as an Emerging Public E-Service

Martin Lnenicka

***Abstract*—**Governments collect and produce large amounts of data. Increasingly, governments worldwide have started to implement open data initiatives and also launch open data portals to enable the release of these data in open and reusable formats. Therefore, a large number of open data repositories, catalogues and portals have been emerging in the world. The greater availability of interoperable and linkable open government data catalyzes secondary use of such data, so they can be used for building useful applications which leverage their value, allow insight, provide access to government services, and support transparency. The efficient development of successful open data portals makes it necessary to evaluate them systematic, in order to understand them better and assess the various types of value they generate, and identify the required improvements for increasing this value. Thus, the attention of this paper is directed particularly to the field of open data portals. The main aim of this paper is to compare the selected open data portals on the national level using content analysis and propose a new evaluation framework, which further improves the quality of these portals. It also establishes a set of considerations for involving businesses and citizens to create e-services and applications that leverage on the datasets available from these portals.

***Keywords*—**Big data, content analysis, criteria comparison, data quality, open data, open data portals, public sector.

I. Introduction

OVER the last years, an increasing number of governments have started to open up their data. The so-called Open government movement has resulted in the launch of numerous open data portals that aim at providing a single point of access for government data [1]. Open government is a new recent phenomenon in which public sector data are made available and can be used by everybody for what it seems an unlimited amount of purposes [2]. By publishing government 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 [3]. Although there are many different sources of data, government data is particularly important because of its scale, breadth, and status as the canonical source of information on a wide range of subjects [4]. As publicly available data can often be generated and provided in huge amounts and through multiple sources, businesses as well as public sector deal with huge quantities and varieties of data on one hand and faster expectations for analysis on the other [5], [6]. These large amounts of data can be very useful for conducting advanced scientific research in the social, political, economic, or management sciences, which can lead to a better understanding of the serious problems that modern societies face. Furthermore, these data have a significant potential for reuse for developing new products and services, possibly in creative combinations with other open data sources [7]. The possibilities to better use available data are growing due to the technical facilities and advancement to merge and analyze different datasets [8]. The emergence of open and big data use and reuse is yet another phase of the ongoing Information and Communication Technologies (ICT) revolution and the public sector is at the center of the current shift to openness [2], [9], [10].

The core idea behind Open Government Data (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. The implementation of OGD takes dedicated and sustained policy attention. Affecting widespread impact through the release of OGD relies not only upon the supply of high-quality data, but also upon the capacity of users to work with these data, and the ability of governments to engage proactively with those users [11]. As a result, during the last years different solutions were developed to support the whole lifecycle of the open and big data reuse such as data discovery, cleaning, integration, processing and visualization [1], [12]. Also, much more work needs to be done in measuring and understanding the impact, value and return on investment of OGD [9], [10], [13].

Large amounts of data and information are daily produced by the European public authorities being the largest single source of information in Europe with an estimated market value of 32 billion Euros [14]. Also [15] estimated that aggregate direct and indirect economic impacts from the use of open and big data across the whole EU28 economy are of the order of billions Euros annually. The resulting economic gains can be put into three broad categories: resource efficiency improvements, product and process improvements and management improvements through evidence based, data-driven decision making. But, disclosing these huge amounts of data does not necessarily equate to more transparency and does not necessarily facilitate accountability [16]. Except the economic importance, there are additional issues concerning the regulation of government data such as discoverability, harvesting, community engagement, and interoperability [4].

In 2012 and again in 2014, the United Nations issued big data and OGD for their E-Government Survey reports, which summarized how governments utilized these data to better serve and protect their people [13], [17]. Also the International

M. Lnenicka is with the Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Studentska 95, 532 10, Pardubice, Czech Republic (phone: +420-466036075; e-mail: martin.lnenicka@gmail.com).

Telecommunication Union [18], the World Economic Forum [19], or Waseda University [20] emphasizes the importance of open and big data for the public sector.

II. Problem Formulation and Research Methodology

OGD change the role of the public sector to the information publisher, which in turn may result in a change of power distribution between the public and private sectors, as well as between the government and the general public, where are chances that the work of the government will improve due to increased participation, collaboration and transparency, which will further strengthen democracy.

An ability to discover the relevant data is a prerequisite to unlocking the potential of open and big data. Creating a portal of available datasets is a way how to make these datasets more accessible and thus easier to find [21]. However, [22] 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 [23] are in agreement with the claim that results of data reuse are not discussed and only little feedback is gained by data providers (public sector), in this way barely supporting policy and decision-making. This raises the question about the reasons inhibiting the interest to innovate using open data. Therefore, to solve this problem, an in-depth comparison and classification of open data portals and their quality, which emphasized the importance of open and big data, should be conducted.

This paper aims to provide an extensive overview of open data portals in the world by using the method of content analysis and analyzing the results of a literature overview, conferences and workshops. The aim of this paper is three-fold. Firstly, a review of related works of open, big and linked data and related open government initiatives will be discussed. Subsequently, the selected open data portals will be evaluated according to the global open data rankings. And finally, an evaluation framework for the future research will be proposed. This framework can be used to improve open data portals as well as open data infrastructures, strategies and initiatives.

Defined problem is solved using the methods of content analysis and comparison, when primarily scholarly articles, books and web resources related to the topic of open, big and linked data, open data portals and catalogues, public sector and methods of comparison and evaluation of related benefits and risks are used.

III. Literature Review and Background

The topic of open and big data is generating interest among practitioners in the public sector as well as in the private sector. Open government acts as an umbrella term for many different ideas and concepts. The definition mostly consists of transparency, participation and collaboration of government towards third actors like the economy or the citizenship. Most often, open government is equated with e-government and the usage of ICT [2], [15], [16], [24].

However, the topic is not entirely new, as the reuse of Public Sector Information (PSI) has been the subject of longer debates and EU directives [14]. 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 reuse in order to remove barriers such as discriminatory practices and monopoly markets by harmonizing the regime for the reuse of PSI [14], [25]. 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 [4]. The number of open data initiatives has grown from two to over three hundred in the period 2009–2014 [8], [25], and the membership in the Open Government Partnership (OGP) has gone from eight in 2011 to sixty-five participating countries in 2015.

A. Open and Linked Data, Benefits, Issues and Challenges

The literature on reuse of open data often circles around their potentials [2] and the economic value of government data, while the literature on open government is in a higher grade directed towards government policy and centered on how use of open data can contribute to the generation of social value in collaborative settings [8], [25]. As mentioned above, interest in the concept of open data has been around for many years [26] 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 [5].

Open data are a piece of content or data if anyone is free to use, reuse, and also redistribute it – 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 [9], [26], [27]; [9] then 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. Cowan, Alencar and Mcgarry [28] used several practical examples in an attempt to illustrate many of the related issues and allied opportunities of open data. Also, 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 [23], [26], [29].

Linked data then describe a method of publishing structured data so that it can be interlinked and become more useful through semantic queries. Linked data are a way of publishing data in such a way that it can facilitate the interaction between different data sources, while the concept of open data is oriented to a freely accessible data and without any restrictions at all to the people [12], [30].

Currently the most promising implementation of linked data is based on Semantic Web philosophy and standard Web technologies but in contrast to the Semantic Web vision, it is about publishing structured data in Resource Description Framework (RDF) using Uniform Resource Identifiers (URIs)

rather than focusing on the ontological level [1], [12]. Linked open data are the combination of both: to structure data and to make it available for others to be reused. Data interlinking practice is highly recommended for lowering technological and cost barriers of data aggregation processes [2], [4], [12]. Kalampokis, Tambouris and Tarabanis [1] then deem that the linked data paradigm must be first adopted for constructing the technical infrastructure that is essential for employing data analytics in a decentralized manner on the Web.

Although open data provide many opportunities and capabilities for the public sector, the publication and use processes of open data are complex and it is not easy to predict how users will use open data, when they will use them, and how they will be used in the future [23]. 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 reuse of these data [13]. Various models of processes in and around open data have been put forward under different headings. They have been termed the open data lifecycle, the open data value chain or plain open data process. According to [23] 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 use open data. At least open data publishers and users are involved, but often many more stakeholders are involved, such as open data facilitators, brokers (e.g. organizations that bring together open data users and producers by providing open data websites), citizens, businesses or open data legislators (e.g. the European Commission and national political parties) [8], [23]. Zuiderwijk and Janssen [23] emphasized the need to apply coordination mechanisms, such as standardization and interconnected processes, due to the complexity, lack of structure, uncertainty, dynamism, and the involvement of varying stakeholders in the open data process. They identified these coordination challenges: inappropriate regulatory environment, fragmentation of open data, unclear boundaries of responsibilities, lack of feedback on and discussion of data use, lack of interconnected processes and lack of standardized and planned processes. Yang and Kankanhalli [22] analyzed the effects of different motivators and inhibitors that influence external stakeholders' willingness to innovate with open data.

The intention of open data publication is to make data available to have them reused by external users, in this way profiting from the wisdom of the crowd, and subsequently to support and improve policy-making and decision-making by discussing data and providing feedback to open data providers [23]. However, pushing data out is not sufficient to create value. Robust engagement models and strategies also need to be in place to allow two-way dialogue to take place between the public sector and the users of government data (e.g. individual citizens, businesses, civil society organizations or academics). It should be weighed out whether, how and which supplied government data can be published [2]. This is key for governments to focus on user need and for users to provide feedback on the datasets they would like to see released as a priority which they consider of greater value or more likely to be used by the community [24].

B. Big Data, Platforms and Analytics

At present, although the importance of big data has been generally recognized, people still have different opinions on its definition. However, big data fundamentally mean data sets that could not be perceived, acquired, managed, and processed by traditional technologies and software/hardware tools within a reasonable time [5]. The changes that are required to face are known as the 3Vs: Velocity (data coming at real-time vs near time or coming in streaming), Volume (data coming in TB transactions, or in tables or in files) and Variety (data coming whether in structured or unstructured way), some other authors also include additional Vs (value, veracity or variability), intended to capitalize on an apparent improvement to the definition of big data [29], [31], [32].

Big data can be also viewed from various perspectives and in various dimensions: organizational, technological, legal and economical [33]. A lifecycle of big data can be distinguished into four phases: data generation, acquisition, storage and analysis [5]. Tien [34] identified four steps or components to big data processing: acquisition, access, analytics and also application. For each step, benefits, potential concerns, impact and selected platforms are mentioned based on the various elements.

Consequently, some of the difficulties related to big data include capture, storage, search, sharing, analytics, and visualizing [6]. It indicates that efficient technologies and platforms together with suitable methods have to be developed and used to analyze and process big data [5], [34]. The main challenges of data-intensive computing are managing and processing exponentially growing data volumes, significantly reducing associated data analysis cycles to support practical applications, and developing new algorithms which can scale to search and process massive amounts of data [6], [34]. Numerous notable attempts have been initiated to exploit massive parallel processing architectures as reported in [5], [6], [29], or [32]–[35].

Big data analytics is about exploring these large volumes of data looking for trends, previously unknown correlations and patterns, which can be used to improve strategic planning, make better decisions, increase profits, etc. [6], [35]. Big data analytics should help to better target customer marketing, improve product analytics, business planning, supply chain management, analysis for fraud, waste and abuse [32]. In general, the growing demand for large-scale data mining and data analysis applications has spurred the development of novel solutions from both the industry (web-data analysis, clickstream analysis, network-monitoring log analysis) and the sciences (analysis of data produced by sensor deployments, high-throughput lab equipment) [6], [36]. The most important capability is advanced analytics to uncover previously hidden patterns. With new types of data comes the need to apply new types of algorithms such as entity analytics, network analytics, text analytics, and real-time scoring. Scalability is important because improved accuracy and trust in your data means your

users are more likely to want to integrate additional data sources or increase data volumes. Analytics must be able to push these algorithm processes to interpret text, images and video streams [32].

Che, Safran and Peng [36] and Chen, Mao and Liu [5] reviewed the state of the art frameworks and platforms for processing and managing big data as well as the efforts expected on big data mining. Singh and Reddy [35] provided an in-depth analysis of different platforms available for performing big data analytics and assessed the advantages and drawbacks of each of these platforms based on various metrics such as scalability, data I/O rate, fault tolerance, real-time processing, data size supported and iterative task support. Also [6] analyzed some of the different analytics methods and tools which can be applied to big data as well as the opportunities provided by the application of big data analytics in various decision domains. They concluded that big data analytics can be applied to leverage business changes and enhance decision making by applying advanced analytic techniques on big data and revealing hidden insights and valuable knowledge. Along similar lines, [33] studied and reviewed the issues, techniques and applications of big data, with an emphasis on future business intelligence architectures. Loshin [32] suggested that as a way to properly ground any initiatives around big data, one initial task would be to evaluate the business's fitness as a combination of the five factors: feasibility, reasonability, value, integrability and sustainability. Demchenko et al. [29] introduced the big data lifecycle management model that includes all the major stages and reflects new challenges and specifics in the big data management.

As the amount and the variety of data is 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. Caballero, Serrano and Piattini [31] focused on the evaluation of data quality. They claimed that more than ever the need for assessing the quality-in-use of big data sets gains importance since the real contribution of a dataset to a business can be only estimated in its context of use. The most important characteristic for assessing the level of quality in use of heterogeneous data sets for big data projects 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 [34]. 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 and uses, the provision of good quality metadata is as important as the quality of data themselves.

C. Open and Big Data in the Context of e-Government

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 initiatives. The early 2010s has added new indices to the e-government development research, which are focusing on the new trends in ICT such as cloud computing, open data, big data, social media, etc. These are e.g. Asia Cloud Computing Association's index, Business Software Alliance Global Cloud Computing Scorecard, Open Data Barometer and Web Index by the World Wide Web Foundation or Open Knowledge Foundation's index.

There are also some frameworks, indices and tools which aim to measure and score open data and openness in the selected countries. The Global Open Data Index assesses the state of open government data around the world and has been developed to help answer such questions by collecting and presenting information on the state of open data around the world to ignite discussions between citizens and governments [37]. This index was firstly introduced in 2013 and covered 60 countries. In 2014 it benchmarks 97 countries by looking at ten key datasets in each country (place): election results, company register, national map, government spending, government budget, transport timetables, legislation, national statistics, postcodes/zipcodes and pollutant emissions. Each dataset in each place is evaluated using nine questions that examine the technical and the legal openness of the dataset. In order to balance between the two aspects, each question is weighted differently and worth a different score. Together, the six technical questions are worth 50 points, the three legal questions are also worth 50 points [37].

The Open Data Barometer report aims to uncover the true prevalence and impact of open data initiatives around the world. It analyses global trends, and provides comparative data on countries and regions via an in-depth methodology combining contextual data, technical assessments and secondary indicators to explore multiple dimensions of open data readiness, implementation and impact [11]. The report scores countries on: readiness to secure benefits from open data, including the legal, political, economic, organizational, social, and technical foundations that can support the supply and use of open data; implementation of open data practice, measured through the availability of data across 15 key categories, and the adoption for those datasets of the common practices set out in the Open Definition and the OGD Principles; impacts of open data, measured through media and academic mentions of data use and impact. Data collection includes peer-reviewed expert surveys, a review of open data laws, datasets available by country, and socioeconomic and political secondary data. The first ranking introduced in the 2013 report covered 77 countries. The second edition then evaluates 86 countries.

IV. Open Data Portals and Related Applications

One of the first problems to be solved when working with any data is where to find it. In using data, one needs exactly the right dataset, i.e. with the right variables, for the right year, the right area, etc., and web search engines, while excellent at finding documents relevant to a given term, do not have enough metadata to find datasets like this, particularly since their main use case is for finding web pages rather than data [4]. Thus, to make open data used, dataset needs to be well described and tools have to be available for reusers. Open data have to be in a good quality for others to transform them into knowledge and make them useful [38]. To solve this problem

of discoverability, in the last few years, an increasing number of governments have set up data portals, specialized websites where a publishing interface allows datasets to be uploaded and equipped with high-quality metadata [4], [24]. The same reasons for opening data in the first place are also valid for open data portals: ensure that the tools are co-created with the reuse community, remove the hurdles to start reusing data and try to collaborate with projects outside of one's organization to leverage the data quality [38].

An open data portal is one of the solutions that should be used to significantly improve discoverability of free available datasets [21]. However, [26] 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. [10] presented seven categories of risks associated with open data: 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 portal at the national level [26]. Thus, it is sensible to argue that different nations have different capabilities in developing and implementing their OGD efforts [39]. Finally, what is needed most is the participation and collaboration of citizens and businesses in using the centralized open data portal [2], [7], [8], [14].

A. Definition and Importance of Open Data Portals

The open data portal is a web-based system used to collect existing data from multiple sources that may be in different formats, and publish these data 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 [3], [4]. Open data portals are the interfaces between government data on one side and reusers on the other side. As interfaces, open data portals must be considered as infrastructures. They enable or restrain actions and define a field of possible uses of released data. This lack of neutrality justifies considering open data portals as political objects, which contribute to the governance of released data. Open data portals generally have catalogues of datasets along with some metadata to describe the institution releasing the dataset as well as content of dataset in addition to geography, jurisdiction and time period of data [40].

Thus, the open data portal is basically a catalogue, which is a collection of catalogue records, and contains metadata for a collection of datasets. 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 [21]. 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 [41]. Each dataset can also comprise several data resources [4]. Open data portals usually feature keyword search and browsing interfaces to help users find relevant datasets and retrieve corresponding metadata. As an alternative to making raw data directly available for download, several projects offer web-based data APIs that enable developers to access data within their applications [42]. A sufficient description of a portal should clearly distinguish themes from keywords, while themes are always chosen from a controlled vocabulary, tags are not [43].

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 resources i.e., files or service end points. Using these metadata, users can quickly find the data they need with searching and filtering features [4]. 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 [41], [43]. Some authors also proposed their own DCAT RDF vocabulary as an interchange format to enable standardized description of data catalogues, e.g. in [43].

Dataset format also needs an immediate attention as it may lead to lot of issues of interoperability and integration [40]. 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 [14], [27], [44]. Data standards, codes, vocabularies and schemas are also important aspects of datasets [40]. Some of the challenges of these datasets can be understood as technical problems addressing information storage, access, inquiry, and display. Another way to understand the challenges are as management problems such as defining the rationale and internal processes of data collection, analysis, management, preservation, and access [44].

Since the launch of the first open data portals by the United States government in 2009 and the United Kingdom in 2010 to provide a single point of access to data from multiple public institutions, 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 [22]. These portals then provide a wide range of information significant to the daily lives of citizens such as transport timetables, local government spending, national map, election results, etc. Examples for the increasing popularity of data portals are OGD portals [26], data portals of international organizations and Non-Governmental Organization (NGO)s, scientific data portals as well as master data catalogues in large businesses [27], [30]. Numerous countries, including a good number of EU Member States, have followed along with some local (e.g. city) governments [4].

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 world-wide entry point for finding and using government data. CKAN enables organizations to upload or link and describe data sources using comprehensive meta-data schemes. The user interface of CKAN has been translated into a wide range of languages, and users can choose the language in which they interact with the site. Similar to digital libraries, networks of such data catalogues can support the description, archiving and discovery of data on the Web [4], [30]. From open data portals users may access datasets generated for application development. In addition, government agencies may also post a challenge online for users to solve [22].

B. Models and Frameworks of Open Data Portals

Alexopoulos et al. [45] 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. Charalabidis, Loukis and Alexopoulos [7] then 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. Value dimensions of their model are organized into three value layers, which correspond to efficiency (capabilities it offers to the users), effectiveness (support of users for achieving their user-level and provider-level objectives) and the value associated with users' future behavior. 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.

Kostovski, Jovanovik and Trajanov [3] developed an open data portal, with the use of the technologies of the Semantic Web. It allows users to publish, manage and consume data in machine-readable formats, interlink their data with data published elsewhere on the Web, publish applications build on top of the data, and interact with other users. Also [42] developed a portal based on Semantic Web principles to support the deployment of linked OGD.

C. Open Data Portals Classification and Comparison

One of the first comparisons of the selected open data portals was conducted by [43]. 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, Pardo and Cook [39] then 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 [40] then compared 30 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. Van der Waal et al. [4] describe the key functionality of open data portals and presented a conceptual model to make data portals the backbone of a distributed global data warehouse for the information society on the Web.

Based on the geographical coverage (administration level), open data portals can be divided into the following groups [21]: 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. Based on the maturity of open data portals, [38] proposed a five stages system to represent the main function or affordance that the data portal is built or used for. The 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 reuse of the datasets. The fourth category of open data portals takes care of the data publication itself. Finally, a data hub is set up where data becomes a common resource.

As local, state, federal and international organizations publish content there are understandable concerns about misinterpretation. Best practices are being formulated for and by the bourgeoning number of data publishers and consumers. Best practices for the open and linked data ecosystem are being defined and include guidance for departments and agencies on procurement, vocabulary selection, versioning, stability, URI construction, conversion from legacy data [27]. As well as the official public and private sector sponsored portals, there are numerous unofficial sources of the open data, usually compiled by citizens, communities or aggregators.

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 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 [45].

Classification of the selected open data portals can be seen in the Table I. It extends the results of [12]. Some authors consider open data aggregators as a basic and the most important category of data catalogues [27]. These aggregators then harvest datasets from national open data portals [4]. Although, the debate about open data is often reduced to

OGD, there are also other type of open data such as open business data (OBD), open citizen data (OCD) and open science data (OSD). Actually, descriptions of some of the analyzed data portals indicated that they are mixed portals containing not only OGD, but also OBD, OCD or OSD.

TABLE I
CLASSIFICATION OF VARIOUS OPEN DATA PORTALS

Category (focus)	Selected open data portals
Digitized data from libraries and e-books	http://arxiv.org/, http://www.lib.powerdata.ir/, https://www.bookshare.org/, https://openlibrary.org/, https://www.gutenberg.org/, etc.
International organizations' data	EU (http://publicdata.eu/), UN (http://data.un.org/), World Bank (http://datacatalogue.worldbank.org/), WHO (http://apps.who.int/gho/data/), http://opendataforafrica.org/, http://www.opendatalatinoamerica.org/, etc.
News data	API's of The New York Times, The Guardian Data Blog, iDnes.cz, etc.
OGD – national open data portals	AU - http://data.gov.au, CA - http://open.canada.ca, DE - https://www.govdata.de/, UK - http://data.gov.uk/, USA - http://www.data.gov/, etc.
Open data aggregators (primarily OGD)	http://datacatalogs.org/, http://knoema.com/, Google Public data explorer, http://opengeocode.org/opendata/, http://opengovernmentdata.org/data/catalogues/, http://datos.fundacionctic.org/sandbox/catalog/faceted/, http://opengovernmentdata.org/data/catalogues/, etc.
OSD (primarily universities)	https://www.opensciencedatacloud.org/publicdata/, https://data.csiro.au/, http://statistics.ucla.edu/, http://sos.noaa.gov/Datasets/, etc.
Social data (primarily OCD and also OBD)	The best place to get social data for an API is the site itself: Instagram, Facebook, Twitter, GetGlue, Foursquare, pretty much all social media sites have their own API's.
Spatial data	http://www.openstreetmap.org/, https://www.sharegeo.ac.uk/, http://nws.noaa.gov/gis/, http://gcmd.nasa.gov/, http://www.iscgm.org/.
Sports data	http://www.pro-football-reference.com/, http://sportsdatabase.com/, http://developer.espn.com/
Weather data	http://www.wunderground.com/, http://www.weatherbase.com/, http://openweathermap.org/, http://ncdc.noaa.gov/cdo-web/datasets, etc.

D. Applications Using Open Data

Applications are an important dimension of open data initiatives. Applications may be developed by governments for easy access to voluminous government data or they may be developed by civil society or developer community to deliver services around government data. Businesses may also develop applications to provide added value and customized services to citizens [40].

This importance appears to be particularly true when data and knowledge from all levels of government, NGOs, citizens and businesses are integrated in applications. There are many examples of the use of open data that could make the society not only more accountable but also more safe and secure from phenomena such as climate change [28]. Authors in [28] then described six examples from different domains, which include environment, land development, indigenous societies and social and community support. These examples are an attempt to illustrate the breadth and complexity of the possible applications for open data to support societal security, societal and government accountability, community development and business.

Typical input data for the development of applications using open data are e.g. lists of all government spending, energy use, complete overview of current work on the roads, air quality, etc. Typical applications are interactive such as the online budget, the map of traffic jams, parking availability (in order to monitor the level of utilization of parking parks and anticipate future needs, rather than for everyday management), etc. Additionally, data used through applications should lead the government or city to show citizens how they fulfill the responsibilities and promises.

Some of the successful implemented business or citizen-driven innovations with government-released open data can be found e.g. in [22]. The official open data portal Data.gov.uk registers 364 apps based on open data, e.g. Numberhood – shows how the local area fares on the important issues: the economy, unemployment, education, health, crime or housing; FixMyStreet – report, view, or discuss local problems like graffiti, fly tipping, broken paving slabs, or street lighting, reports are then sent to the local council; BUSit London (UK) – uses London Bus data from Transport for London to plan your multi-leg bus journey in the capital. Other applications worldwide are e.g.:

- Dunny Directories (Australia) – a location based mobile application, which provides the ability to easily locate public toilets throughout Australia.
- mojePaństwo (Poland) – engages citizens with information about what is happening in their parliament.
- LiveTraffic (Singapore) – accesses data from various government sources to provide customized real-time navigation for drivers.
- ShowNearBy (Singapore) – a location-based service delivering business intelligence solutions for Singapore's private, people and public sectors based on public information from the government.
- CrimeReports (USA) – offers a family of affordable, easy-to-use software tools for law enforcement agencies to understand crime trends and share current neighborhood crime data with the public.
- Park It DC (USA) – allows user to check a specific area in the district capital for parking information.

TABLE II
COMPARISON OF THE SELECTED OPEN DATA PORTALS

Country	Open data portal's URL	English version	CKAN powered	Number of datasets	Global Open Data Index 2014		Open Data Barometer 2014	
					Rank	Value	Rank	Value
Argentina	http://datospublicos.gov.ar/	N	Y	26	48	42	37	35.7
Australia	http://data.gov.au/	Y	Y	5254	5	72	10	68.3
Austria	https://www.data.gv.at/	N	N	1548	22	59	16	58.5
Bahrain	http://www.data.gov.bh/	Y	N	N/A	N/A	N/A	62	15.4
Belgium	http://data.gov.be	Y	N	117	53	39	27	47.3
Bermuda	http://bermuda.io/	Y	Y	16	63	34	N/A	N/A
Brazil	http://dados.gov.br/	N	Y	469	26	54	21	52.1
Burkina Faso	http://data.gov.bf/	N	N	141	59	36	74	11.3
Canada	http://open.canada.ca/	Y	N	217031	22	59	8	74.5
Colombia	http://datosabiertoscolombia.cloudapp.net/	N	N	998	12	66	40	32.4
Costa Rica	http://datosabiertos.gob.go.cr/	N	N	250	54	38	41	31.3
Cyprus	http://www.data.gov.cy/	Y	N	139	87	21	N/A	N/A
Czech Republic	http://cz.ckan.net/	Y	Y	180	12	66	17	58.1
Denmark	http://data.digitaliser.dk/	Y	N	757	2	83	9	70.1
Ecuador	http://datosabiertos.ec/	N	N	82	43	44	38	35
El Salvador	http://www.datoselsalvador.org/	N	N	71	57	37	N/A	N/A
Estonia	https://opendata.riik.ee/	N	Y	5	N/A	N/A	13	60.2
Finland	http://data.suomi.fi/	Y	N	N/A	4	73	12	66.5
France	https://www.data.gouv.fr/	N	N	14027	3	80	4	80.2
Germany	https://www.govdata.de/	N	N	9959	9	69	11	67.6
Ghana	http://data.gov.gh/	Y	N	1470	82	24	48	28
Greece	http://data.gov.gr/	Y	N	75	54	38	31	40.8
Hong Kong	http://www.gov.hk/en/theme/psi/datasets/	Y	N	56	54	38	N/A	N/A
Hungary	http://opendata.hu/	Y	Y	48	36	48	33	38.3
Chile	http://datos.gob.cl/	N	N	1187	19	61	15	58.7
China	http://portal.opendatachina.com/	Y	Y	6	57	37	46	28.1
Iceland	http://opingogn.is/	N	Y	23	16	64	28	46.6
India	https://data.gov.in/	Y	N	13156	10	68	39	33.2
Indonesia	http://data.go.id/	N	Y	940	45	43	36	36.2
Ireland	http://data.gov.ie/	Y	Y	509	36	48	32	40.7
Israel	http://data.gov.il/	N	N	N/A	40	46	20	53
Italy	http://www.dati.gov.it/	N	N	11048	25	55	23	50.6
Japan	http://www.data.go.jp/	Y	Y	12807	19	61	19	53.6
Kazakhstan	http://data.egov.kz/	Y	N	243	N/A	N/A	50	25.9
Kenya	https://www.opendata.go.ke/	Y	N	622	85	22	51	25.8
Latvia	http://data.opendata.lv/	Y	N	10	34	51	N/A	N/A
Lithuania	http://opendata.gov.lt/	N	N	282	70	32	N/A	N/A
Luxembourg	http://www.opendata.lu/	N	N	6	N/A	N/A	N/A	N/A
Malaysia	http://data.gov.my/	Y	N	117	N/A	N/A	43	30.8
Malta	http://opendatamalta.org/	Y	Y	9	31	52	N/A	N/A
Mexico	http://datos.gob.mx/	N	Y	297	28	53	24	50.1
Moldova	http://data.gov.md/	Y	N	793	43	44	N/A	N/A
Morocco	http://data.gov.ma/	N	N	105	79	25	56	21.1
Nepal	http://data.opennepal.net/	Y	N	201	63	34	63	14.6
Netherlands	https://data.overheid.nl/	Y	N	3250	16	64	6	75.8
New Zealand	https://data.govt.nz/	Y	N	2813	5	72	5	80
Norway	http://data.norge.no/	N	N	602	7	71	7	74.6
Pakistan	http://data.org.pk/	Y	Y	53	41	45	67	12.6
Philippines	http://data.gov.ph/	Y	Y	660	71	31	53	23.2
Poland	http://pl.ckan.net/	Y	Y	111	48	42	35	37
Portugal	http://www.dados.gov.pt/	Y	N	621	39	47	29	46.1
Romania	http://data.gov.ro/	Y	Y	204	16	64	N/A	N/A
Russia	http://data.gov.ru/	Y	N	2456	45	43	26	48.3
Serbia	http://rs.ckan.net/	Y	Y	168	48	42	N/A	N/A
Singapore	http://data.gov.sg/	Y	N	8315	63	34	30	46.1
Slovakia	http://data.gov.sk/	Y	Y	221	61	35	N/A	N/A
Slovenia	http://si.ckan.net/	Y	Y	36	26	54	N/A	N/A
South Korea	https://www.data.go.kr/	Y	N	11915	28	53	18	57.7
Spain	http://datos.gob.es/	N	N	7742	31	52	14	59.9
Sweden	http://oppnadata.se/	Y	Y	54	12	66	3	83.7
Switzerland	http://opendata.admin.ch/	Y	N	1849	24	58	22	51.3
Taiwan	http://data.gov.tw	N	N	N/A	11	67	N/A	N/A
Tunisia	http://data.gov.tn/	N	N	177	63	34	45	28.6
Uganda	http://www.data.ug/	Y	Y	374	N/A	N/A	64	14.5
United Kingdom	http://data.gov.uk/	Y	Y	23380	1	97	1	100
United States	http://data.gov/	Y	N	138584	8	70	2	92.7
Uruguay	https://catalogueodatos.gub.uy/	N	Y	106	12	66	25	49.4

V. Case Study

Only open data portals on the national level are evaluated, no 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 Global Open Data Index and Open Data Barometer from 2014. Together, they cover 117 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 datacatalogs.org, opengeocode.org/opendata/ and ckan.org/instances/; and the identified portal's URL is opened to examine whether it is in working condition.

As a result, in this first step 50 countries were omitted from the comparison, because they have no open data portal on the national level dedicated exclusively to publishing OGD. Most of these countries have only statistical institute or office portal and don't develop open data portal on the national level. The results presented in Table II are based on the content analysis of 67 open data portals on the national level conducted in February 2015. Since the purpose of this case study is to choose the most appropriate portals for the future comparison using the new evaluation framework, only three criteria were defined. The first one is the existence of the English version of the portal, because the evaluation will be carried out only in English. The second one is the data management system of the portal. In this case it is CKAN, which is the most widely used open-source data portal platform [30]. The third criterion evaluates the size of portal and the number of accessible datasets.

The results show that 42 open data portals on the national level are accessible in English language. Also 25 portals are powered by CKAN. Fig. 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, France, the United States and United Kingdom open up more of their datasets to the public. These results are in agreement with the rankings of the Global Open Data Index and Open Data Barometer. In the Global Open Data Index's rank order, the highest level of openness exists in the United Kingdom, Denmark and France. In the Open Data Barometer's 2014 rank order, the highest level of openness exists in the United Kingdom, in the United States and Sweden.

The next step will be the comparison of the selected portals using the evaluation framework, which can be seen in the Table III. This framework is based on the in-depth literature review and author's experiences and knowledge gained in the first step of the open data portals comparison. It is divided into two parts. The first one focuses on the general characteristics that consist of technical perspective, availability and access, and communication and interaction. The second one evaluates the general characteristics of datasets. 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 [3], [7], [11], [21], [24], [37], [43].

TABLE III
An Evaluation Framework for Open Data Portals

General characteristics of open data portals	
List of criteria	*Description*
1. Technical perspective	
1.1 Authority and responsibility	There has to be information about the authority, which hosts the portal and the governance model/institutional framework supporting data provision models [24].
1.2 Data management system	There has to be information about the data management system, which is used to power the portal [3], [43].
1.3 Language	The portal should offer more language versions to gain more users (attention) and improve the overall quality of this portal [7].
1.4 Free of charge	All datasets and services have to be available free of charge and without any restrictions [11], [37].
2. Availability and access	
2.1 Number of datasets	The size of portal refers to the number of datasets it includes [3], [43].
2.2 Number of applications	Number of applications developed based on the open data published [24].
2.3 Search engine	The portal should provide strong dataset search capabilities using different criteria [7], [21], [24], [43].
2.4 API	The portal should provide API for stakeholders to develop applications using open data [7], [24], [43].
2.5 User account	The portal should support user account creation in order to personalize views and information shown [7].
2.6 Categories	The portal should clearly distinguish categories (themes) from tags (keywords) [43]. Same tags should be used to classify data of the same type and category [7], [21].
2.7 Tags	
3. Communication and interaction	
3.1 Forum	The portal should provide an opportunity to provide feedback on the data and forum to discuss and exchange ideas among the stakeholders [7], [24].
3.2 Request form	The portal should provide a form to request new type or format type of open data [21], [43].
3.3 Help	The portal should include high quality of documentation and help functionality to learn how to use the portal [7], [24].
3.4 FAQ	The portal should provide a FAQ section [24].
3.5 Social media	Is the portal connected to a social media platform [24]?
General characteristics of datasets	
1. Title and description	Open data should be provided together with their description and also how and for what purpose they were collected [7], [21], [37].
2. Publisher	Open data should be provided together with their publisher to verify authenticity of their source [43].
3. Release date and up to date	All information in the dataset should be up to date [11], [21], [37], [43].
4. License	Data that doesn't explicitly have an open license are not open data [11], [37], [43].
5. Geographic coverage	It should be determined if the coverage of open data is on the national, regional or local level [43].
6. Dataset URL	Dataset URL should be available [43].
7. Dataset size	Dataset size should be available [3], [43].
8. Number of views	Total number of online views for a dataset [24].
9. Number of downloads	Total number of downloads for a dataset [24].
10. Machine-readable formats	Open data should be provided in formats that are as convenient, easy to analyze and modifiable as downloadable files in well-known formats [7], [11], [21], [24], [37], [43].
11. User rating	The portal should allow to collect user ratings and comments on a dataset [7].

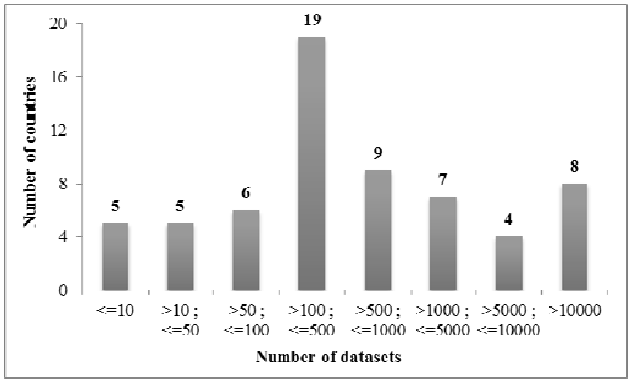

Fig. 1 Histogram of numbers of available datasets

Each criterion will be converted to a question to be included in a questionnaire to be distributed to users. These questions will be then evaluated on a five point Likert scale to measure agreement or disagreement with (positive or negative response to) such a statement with (1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree). Only open data portals accessible in English language with more than 100 datasets will be evaluated to showcase the applicability of the proposed framework, i.e. 47 countries. The results will be presented in the future paper.

VI. Conclusion and Future Research Directions

Open and big data can be used to create new business models or increase efficiency of existing ones. Public sector institutions can use insights gained by third parties to improve their own operations. However, it is important to remember about the limits of what open data can achieve by themselves.

This paper presents a preliminary exploration of the status of open data portals worldwide as well as in-depth review of the issues, challenges and opportunities associated with these portals. There were also mentioned several observations in terms of policy and practical implication of open data portals development drawn from the presented comparison such as metadata quality. However, the number of datasets online and the sophistication of open data portals differ. In particular, Canada, the United States and also United Kingdom have published many datasets and launched advanced websites. The new proposed framework can be used to improve open data portals as well as open data infrastructures, strategies and initiatives. The results also showed that open data portals are very promising public e-service with a great potential.

Future work will be dedicated to the use of the proposed evaluation framework and study of various types of datasets and services developed and the way in which it is possible to optimize the return on investment of open data initiatives by selecting relevant datasets and understanding the process by which successful services can be built on top of those datasets.

References

[1] 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.

[2] C. P. Geiger, and J. von Lucke, "Open Government and (Linked) (Open) (Government) (Data) ", *eJournal of eDemocracy & Open Government*, vol. 4, no. 2, 2012, pp. 265–278.

[3] 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.

[4] S. van der Waal, et al., "Lifting Open Data Portals to the Data Web", in *Linked Open Data – Creating Knowledge Out of Interlinked Data,* Springer, International Publishing, 2014, pp. 175–195.

[5] M. Chen, S. Mao, and Y. Liu, "Big Data: A Survey", *Mobile Networks and Applications*, vol. 19, no. 2, 2014, pp. 171–209.

[6] N. Elgendy, and A. Elragal, "Big Data Analytics: A Literature Review Paper", in *Advances in Data Mining: Applications and Theoretical Aspects*, Springer, Heidelberg, 2014, pp. 214–227.

[7] 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, 2014, pp. 2114–2126.

[8] 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.

[9] J. Kucera, and D. Chlapek, "Benefits and Risks of Open Government Data", *Journal of Systems Integration*, vol. 5, no. 1, 2014, pp. 30–41.

[10] 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, 2013, pp. 301–309.

[11] World Wide Web Foundation. Open Data Barometer Global Report (online). 2015. Available at: http://opendatabarometer.org/assets/downloads/Open%20Data%20Barometer%20-%20Global%20Report%20-%202nd%20Edition%20-%20PRINT.pdf (cited 2015-02-15).

[12] 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, 2011, pp. 1–136.

[13] United Nations, United Nations E-government Survey 2014: E-Government for the Future We Want, New York: UN Publishing Section, 2014.

[14] 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*, Springer, Berlin, 2013, pp. 167–180.

[15] S. Buchholtz, M. Bukowski, and A. Śniegocki, Big and open data in Europe: A growth engine or a missed opportunity, 1st ed. Varšava: demosEUROPA, 2014.

[16] R. Lourenço, "Open Government Portals Assessment: A Transparency for Accountability Perspective", in *Proceedings of the 12th IFIP WG 8.5 International Conference: EGOV 2013*, Springer, Heidelberg, 2013, pp. 62–74.

[17] United Nations, United Nations E-government Survey 2012: E-Government for the People, New York: UN Publishing Section, 2012.

[18] International Telecommunication Union, Measuring the Information Society Report 2014, Geneva: Telecommunication Development Bureau, 2014.

[19] World Economic Forum, The Global Information Technology Report 2014: Rewards and Risks of Big Data. Geneva: SRO-Kundig, 2014.

[20] Waseda University. WASEDA - IAC 10th International E-Government Ranking 2014 (online). 2014. Available at: http://www.e-gov.waseda.ac.jp/pdf/2014_E-Gov_Press_Release.pdf (cited 2015-02-15).

[21] 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*, Springer, Berlin, 2013, pp. 152–166.

[22] 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*, Springer, Berlin, 2013, pp. 644–651.

[23] 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.

[24] B. Ubaldi, "Open Government Data: Towards Empirical Analysis of

Open Government Data Initiatives", *OECD Working Papers on Public Governance*, No. 22, OECD Publishing, 2013.
[25] 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.
[26] 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, 2012, pp. 258–268.
[27] B. Hyland, and D. Wood, "The Joy of Data-A Cookbook for Publishing Linked Government Data on the Web", in *Linking Government Data*, Springer, New York, 2011, pp. 3–26.
[28] D. Cowan, P. Alencar, and F. Mcgarry, "Perspectives on Open Data: Issues and Opportunties", in *Proceedings of the 2014 IEEE International Conference on Software Science, Technology and Engineering*, IEEE, 2014, pp. 24–33.
[29] Y. Demchenko, et al., "Addressing Big Data Issues in Scientific Data Infrastructure", in *Proceedings of International Conference on Collaboration Technologies and Systems*, IEEE, 2013, pp. 48–55.
[30] I. Ermilov, et al., "Linked Open Data Statistics: Collection and Exploitation", in *Knowledge Engineering and the Semantic Web*. Springer, Berlin Heidelberg, 2013, pp. 242–249.
[31] I. Caballero, M. Serrano, and M. Piattini, "A Data Quality in Use Model for Big Data", in *Advances in Conceptual Modeling*, Springer, Heidelberg, 2014, pp. 65–74.
[32] D. Loshin, Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph, 1st ed. Elsevier Inc.: Waltham, 2013.
[33] G. Vossen, "Big data as the new enabler in business and other intelligence", *Vietnam Journal of Computer Science*, vol. 1, no. 1, 2014, pp. 3–14.
[34] J. M. Tien, "Big Data: Unleashing Information", *Journal of Systems Science and Systems Engineering*, vol. 22, no. 2, 2013, pp. 127–151.
[35] D. Singh, and C. K. Reddy, "A survey on platforms for big data analytics", *Journal of Big Data*, vol. 1, no. 8, 2014, pp. 1–20.
[36] D. Che, M. Safran, and Z. Peng, "From Big Data to Big Data Mining: Challenges, Issues, and Opportunities", in *Database Systems for Advanced Applications,* Springer, Heidelberg, 2013, pp. 1–15.
[37] Methodology. *Global Open Data Index by Open Knowledge* [online]. 2014. Available at: http://index.okfn.org/methodology/ [cit. 2015-02-15].
[38] 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.
[39] 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.
[40] 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, 2012, pp. 207–216.
[41] R. Cyganiak, and F. Maali, Use Cases and Requirements for the Data Catalog Vocabulary, (Online). 2015. Available at: https://dvcs.w3.org/hg/gld/raw-file/default/dcat-ucr/index.html [cited 2015-02-02].
[42] T. Lebo, et al., "Producing and Using Linked Open Government Data in the TWC LOGD Portal", in *Linking Government Data*, Springer, New York, 2011, pp. 51–72.
[43] 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.
[44] 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.
[45] 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*, Springer, Heidelberg, 2014, pp. 230-241.