*Online Journal of Applied Knowledge Management*
A Publication of the International Institute for Applied Knowledge Management

*Volume 6, Issue 2, 2018*
*Special Issue on Knowledge Management: Research, Organization, and Applied Innovation*

# Open government data, the case of Polish public sector

**Jędrzej Wieczorkowski,** Warsaw School of Economics, Poland, jedrzej.wieczorkowski@sgh.waw.pl

**Ilona Pawełoszek,** Częstochowa University of Technology, Poland, ilona.paweloszek@wz.pcz.pl

## Abstract

*This paper presents the idea of open government data along with the benefits and threats resulting from using open data. We describe the results of our research study on availability of the open data on the example of Poland with particular emphasis on Central Repository for Public Information (CRPI). The comparison of CRPI in Poland and other countries has been discussed. The review of accessible public information has been made with particular focus on data formats. Data formats are an important aspect of open data as they facilitate or impede the reuse of data. The insights from our participant observation in the projects of computerization of public administration are also presented. Although the Open Government Data (OGD) movement can provide a number of benefits, recent study has shown that in Poland it has not achieved its full potential yet.*

**Keywords**: Open data, open government data, linked data, central repository for public information.

## Introduction

### Idea of Open Government Data

A concept of universal right of access to information is not new. It has been a subject of the public debate from the middle of the 18th century (Mendel, 2003). In the second part of 20th century the law of free access to information became a common standard in many countries. According to the current Constitution of the Republic of Poland (1997) (article.61) a citizen shall have the right to obtain information on the activities of organizations of public authority as well as persons discharging public functions. The right to obtain the access to documents and data gives the possibility to use them in various ways. The purposes of using data can be private, social, and commercial.

In the past few years, activities of public administration have been increasingly influenced by digital technology and possibility of enforcing the society's right to information. The changes not only influenced the workflow of public administration and internal communication, but also first of all enabled the interaction between the government and citizens. There is an open debate on the key issues of: (a) to what extent the information gathered by public institutions should be commonly accessible; and, (b) to what degree legal regulations should allow to reuse the information generated from public funds (Hamilton & Saunderson, 2017). These two key issues are entering a wider discussion on the concept of data openness and knowledge built on the basis of private funds. The openness is related with the possibility of reuse of knowledge created by linking different datasets with the aim to obtain new innovative results.

The concept of Open Data describes the datasets, which can be accessed, used, processed and published by anyone, without restrictions of copyright and patent law, with only the requirement to indicate the source of data or allowing for further distribution of the processed content under the same conditions (Kozierski, Kabaciński, Lis, & Kaczmarek, 2013). The idea of Open Government Data (OGD) has emerged at the intersection of the Open Government and Open Data. Its purpose is to publicize information resources created by or on the order of public administration, as well as the free use and distribution of open data by each citizen (Papińska-Kacperek & Polańska, 2015). With this purpose in mind, public administration bodies in many countries started to publish government data on their websites as web services or by Representational State Transfer (REST) interface. The aim is to gain better understanding of public administration policies by citizens, making the administration more effective and trustworthy. Several case studies show positive impact of open data on business, economic growth, prosperity, and innovation (Stagars, 2016; Kitsios, Papachristos, & Kamariotou, 2017; Janssen, Charalabidis, & Zuiderwijk, 2012; Lakomaa, & Kallberg, 2013).

The published data pertain to many disciplines, of which the most popular are: environmental protection, data and statistics on employment, budgets of government bodies, maps, timetables of public transport, etc. Moreover, individual administration bodies or other providers can publish it in a centralized and standardized manner. These repositories can differ in terms of scope, size, and standards of the data published. Undoubtedly, the most important feature of open data sources is their potential usability for concrete tasks. To add value to the development of open society, the platforms of open data should meet certain legal, administrative, and technical requirements. Thus, the question we pose is: how the ideas mentioned above are implemented in practice, particularly how the data openness looks in the Polish public sector? Research shows considerable reluctance of Polish public sector to providing information upon request despite a statutory obligation (Malinowska-Misiąg, 2016). However, the idea of open data means more than a mere passive transparency, which means providing information on request. The active form of informing is also important and it consists of publishing information using the methods most appropriate to the needs and possibilities.

## Research Questions and Methods

The aim of this paper is the analysis and the assessment of availability and usability of the open government data in Poland compared with leading open government countries. This study particularly focused on Central Repository for Public Information (CRPI), which is the main access gateway to open government information in Poland. On the basis of the review the research questions have been posed:

- How to assess OGD in Poland?
- How Poland compares to other countries?

To get an idea of the development of OGD in selected countries, we got acquainted with reports, statistics and articles published by independent institutions.

In order to select the countries to be examined, we decided to treat reports containing OGD rankings as a starting point. We wanted to include in particular OGD and CRPI leaders in their comparison. Such documents were searched on the Internet from February 2017 to December

2018 using Google search engine, which is the search engine market leader according to Net MarketShare ranking (https://www.netmarketshare.com/, 2018). The first key phrase used was “Open Data global report”, the second key phrase used was “OECD open data”. The Organization for Economic Co-operation and Development (OECD) is a reliable source of comparable statistics and economic and social data. A list of websites that have been analyzed is presented in Table 1.

**Table 1.** List of Analyzed Websites

| Site Name | URL | Focus |
|---|---|---|
| Open Data Barometer | https://opendatabarometer.org/3rdedition/report/ | Open government data |
| Global Open Data Index | https://index.okfn.org/ | Open government data |
| Open Data research perspective | https://www.elsevier.com/about/open-science/research-data/open-data-report | Science and research |
| The State of Open Data 2017 | https://www.digital-science.com/resources/portfolio-reports/state-open-data-2017/ | Science and research |
| Global open data | https://www.tomforth.co.uk/globalopendata/ | Data formats |
| Open Data Institute | https://theodi.org/knowledge-opinion/reports/ | Thematic reports |
| The GovLab Index: Open Data – 2016 Edition | http://thegovlab.org/govlab-index-on-open-data-2016-edition/ | Value and Impact of Open Data |
| Transparency international | https://www.transparency.org/news/feature/open_data_promise_but_not_enough_progress_from_g20_countries | Anti-corruption |
| Government at a Glance 2017 | https://www.oecd-ilibrary.org/governance/government-at-a-glance-2017/open-useful-reusable-government-data-index-ourdata-2017_gov_glance-2017-graph139-en | Open government data |

Ultimately the analyzed publications came from OECD (2015), Open Data Barometer Portals (Web Foundation, 2015), European Commission (2016) and Open Government Partnership (https://www.opengovpartnership.org/). The selected publications focus on general aspects of open government data and provide reliable statistics from many countries.

The second part of the study was focused on the content of open data publishing websites (CRPIs) of selected countries (USA, UK, Germany, & Poland). We carried out our research based on exploration of CRPI in Poland and open data portals in other countries from March 2017 to January 2018. As a reference point for the evaluation of the Polish website, the United States of America (USA) website has been adopted, where the largest number of data sets is currently available, moreover, the reports analyzed above often indicate the USA as the world leader in open data. In addition to the USA, we decided to compare to two European countries, which are economically and politically closer to Poland, thanks to the common (so far) membership in the European Union. The United Kingdom (UK) was chosen because this country according to the surveyed reports is most often treated as a leader of CRPI in Europe, as well as Germany, as the largest and the most economically strong state of the European Union, also a direct neighbor of Poland.

This research was carried out by entering individual websites of CRPI portals and analyzing the availability of various data. The review of accessible public information has been made particularly focusing on data formats, which determine the availability and reuse of the data. The aim of this study was to compare the maturity of OGD in Poland to other countries. On the base of the comparison and the study of best practices, recommendations are proposed on how to improve accessibility and reuse of the data. Additionally, to better understand the drivers and obstacles in OGD development, participant observation was conducted by the first author during the projects of computerization of public administration in Poland. Participant observation is a qualitative research method in which a researcher takes part in the daily work of a group of people. It helps to learn explicit and tacit aspects of their activities (DeWalt &, DeWalt, 2011). This method can be used in a variety of disciplines as a tool for collecting data about people and processes (Kawulich, 2005). It is an inductive form of research that can be very helpful in the generation and modification of hypotheses (Krishnaswamy, Sivakumar, & Mathirajan, 2009). In this case participant observation was used while formulating hypotheses for further research and to get insight into the possibilities and social needs related to OGD. The starting point for participant observation was the hypothesis that there is a low awareness of the need and benefits of data sharing in Polish public administration institutions.

In the period from June 2016 to January 2018, the participant observant was involved in the process of defining the assumptions of programs aimed at financing (from European Union & governmental funds) Information and Communication Technology (ICT) projects, aimed in particular at providing access to the OGD. The participant observant also dealt with the assessment of applications for co-financing in such programs. These activities took place as part of the Operational Program Digital Poland (OPDP) in the e-government and open government axis. The program’s goal is, among others, to provide public sector information from administrative sources and scientific resources, as well as improving the possibilities of their re-use. For example, they finance the implementation or extension of services provided by electronic applications using the content of open public-sector information resources or existing public e-services. The resources that are in the possession of public administration and science are also digitized in order to make them widely available and re-used (See https://cppc.gov.pl/). Also, the participant observant took part in the discussions of experts assessing the applications of OGD implementation projects in cooperation with the Digital Poland Project Center, (the government organization overseeing the OPDP). He also participated in the Delphi research and subsequent panels of experts aimed at developing the principles of the new OPDP subroutine supporting financially such projects, for the order of the Ministry of Development. During observation, the participant observant had studied a series of cases related to the tasks of data collecting and publishing. He participated in expert discussions and meetings with representatives of institutions requesting the acceptance of proposed ICT projects in the public sector. The collected information in the form of applications for co-financing the project, personal notes, expert assessment cards, and various documentation regarding the discussed projects were used to understand the OGD problems in Poland for the purposes of this article.

In the research method adopted, it is important to ensure objectivity in the process of observation, data collection and drawing conclusions. The observation was related to the participation in the teams of experts, and in this case, it is important to minimize the inference of improper suggestion with previous own or someone else’s beliefs, in particular related to the confirmation bias phenomenon. The quality of the research procedure in the discussed case is influenced by the procedures used by the expert teams, as the functioning of these teams is subject to observation. The procedures used to evaluate the above-mentioned projects try to minimize the subjectivism of the evaluators. In particular, applications are evaluated by several experts drawn from a larger group of people with appropriate qualifications and experience. Each expert independently completes a very detailed application evaluation sheet taking into account many criteria. At this stage, experts do not have contact with each other. Then they have the opportunity to get to know each other's assessments and, if necessary, a meeting takes place for the purpose of a personal exchange of views, in some cases also with representatives of applicants. In addition to documentation of evaluated projects, detailed documentation is created in the form of subsequent versions of project evaluation sheets made by individual experts. Such documentation in connection with notes made by the participant observant becomes the basis for further inference in the research process. The procedures for the assessment of applications and, as a consequence, the research procedures used, are aimed, on the one hand, to limit the mutual suggestion of the assessments of other experts, and, on the other hand, allow them to discuss and expand their horizons.

The ideas of new projects submitted by various institutions (public administration, science, & culture) were subject to observation. Institutions submitting projects were aware of their own OGD. However, their previous procedures were often based on solutions created in the era of paper-based workflow, and the resources (e.g. scientific studies, cultural works) required digitization. The question is whether the implementation of projects takes full advantage of modern technological capabilities, and is not just an uncritical digitalization of existing processes. In particular, the question is not only which data should be made public, but also whether the data should be made available in an optimal form from the user's point of view. The observations carried out are to help answer the above questions using a practical point of view - based on actual implemented or only requested OGD sharing projects.

# Theoretical Background

## Open Government and Open Data

In various academic literature resources there is an assumption that openness is the fundamental feature of e-government operating in the spirit of democracy of 21th century (Banisar, 2005). OECD defined the features of the open government in the following way: transparency of government activities, accessibility of government services and information, and reaction on new ideas, requirements and needs. These trends are viewed as the base for other advantages for government and society, such as: creating a knowledge base to establish the policy, enhancing the integrity, discouraging corruption, and building social trust (Curristine, & Abbott, 2005).

All the aforementioned tasks rely on information policy of the government. Therefore, open government requires implementation of the OGD, which is the data produced or commissioned by government or government controlled entities and it can be freely used, reused and redistributed by anyone. According to the Open Knowledge Foundation (https://opengovernmentdata.org/) the data of the public sector should be open for the following reasons:

- ***Transparency:*** Citizens need to know what their government is doing. Transparency isn't just about access, it is also about sharing and reuse.
- ***Releasing social and commercial value:*** By opening up data, government can help drive the creation of innovative business and services that deliver social and commercial value.

- ***Participatory Governance:*** Citizens are enabled to be much more directly informed and involved in decision-making. This is more than transparency.

Comparing the situation of government to other organizations in the context of open data, it is worth to note that it is the government that is obligated to make the data available. Government bodies are acting on the base of public funds. Moreover, the government is privileged because it has the possibility to make other bodies (mainly commercial) provide data. In return for this obligation, the government has to make the data available for citizens and private sector (Mayer-Schonberger & Cukier, 2013). Furthermore, creation and development of new business models on the basis of processing of public data act in the general economic interest. This is particularly important while the possibilities of modern Information Technology (IT) are increasing, especially in the field of processing large data sets. Data mining based on such collections has become a well-established part of modern business, and systems based on this approach play an analytical role as well as being part of operational processes (Szupiluk, 2013).

## Technical Aspects of Open Data and Linked Data

The two basic concepts for efficient publishing of the data are: (a) machine readable format; and (b) open license allowing for using the data. At the present, in many countries the legal acts regarding free access to data contain a clause about the data format. For example, in the USA in the Open Government directive issued on December 2009 can be found:

> “To the extent practicable and subject to valid restrictions, agencies should publish information online in an open format that can be retrieved, downloaded, indexed, and searched by commonly used web search applications. An open format is one that is platform independent, machine readable, and made available to the public without restrictions that would impede the reuse of that information”. (Orszag, 2009, p. 2)

Currently, the data are most often published in a way that requires to enter the website of some publishing institution to find the dataset of interest. The real advantage of publishing open data comes from the possibility of integration of different resources. Integration of data from several sources is complicated and time consuming. The problem is present because each organization has different policy regarding data publishing and the resources are heterogeneous regarding their format. The integration requires easy localization and aggregation of data pertaining to a particular object or phenomenon. It is especially visible when the data are coming from different sources and present the analyzed issue in various contexts.

Integration of open data cannot be performed directly because they have different characteristics. Even if the data describe the same issue, sometimes it cannot be easily evidenced from screening the spreadsheets. To determine the real meaning of data, it is required to analyze not only their content, but also the accompanying metadata that describe the context. Additionally, the diversity of measures, scales, and nomenclature should be taken into account.

Making the data available to the society requires a planning already at the stage of data gathering. The data should be gathered in a way that facilitates their adaptation to open formats and publishing in a form that could be understood by people as well as processed by machines.

The term “Linked Data” refers to a set of best practices for publishing and interlinking structured data on the web for access by both humans and machines (W3C, 2014). The mentioned best practices increasingly often implemented by data providers, allow for creating global data space containing millions of hyperlinks connecting the data that are semantically related. Such a collection of interlinked information is often referred to as the "Web of Data" (Polleres, Amato, Arenas, & Handschuh, 2011).

Berners-Lee outlined a set of guidelines for publishing data on the web in a way ensuring participation in the global data space (Bizer, Heath, & Berners-Lee, 2009):

- Use URLs (Uniform Resource Locator) as names for things
- Use Hypertext Transfer Protocol (HTTP) URL so that people can look up those names
- When someone looks up a URL, provide useful information, using the standards - Resource Description Framework (RDF), SPARQL Protocol And RDF Query Language (SPARQL)
- Include links to other URLs, so that they can discover more things.

The above principles constitute the technical basis for publishing Linked Open Data (LOD), it is notable that using standards is crucial. Considering the different initiatives of publishing Open Data (such as data.gov) it can be seen that there is no consensus as to in what form the data should be published. Two approaches are used: publishing "raw" data, publishing data that underwent statistical processing. Publishing the raw data (in the form of tables from databases) is less expensive for the publisher. Such a form offers extensive analytical possibilities, but it requires knowledge and skills on how to use statistical tools and process the data. Publishing raw data may have undesirable effects (Meijer, 2009). The data taken out of context, having structured numerical form can turn public attention to narrow and not necessarily important (however quantifiable) phenomena. Publishing data after statistical processing (in the form of reports, charts, clearly described, & interpreted factors), provides the users with easy to understand information. However, there is a risk that the reports and statistics can be biased, some trends may be overestimated or underestimated. There is also a threat of manipulation of numbers or graphics to create a false impression.

LOD can be considered in two contexts – data publication and usage. Berners-Lee (https://www.w3.org/DesignIssues/LinkedData.html) suggested a five-star LOD deployment scheme (Table 2), which can also be applied to develop open government data. The 5-five-star model presents the characteristics of practical solutions in the area of open data regarding the level of LOD maturity. Each of the generations of open data development brings some benefits and creates some costs as well for the publishing bodies as for the data consumers.

**Table 2.** The Plan of Implementation of Open Data (Bizer, Heath, & Berners-Lee, 2009; Wood, 2011)

| Generation | Description | Example |
|---|---|---|
| ★ | Publication on the Web, open license | PDF document containing tables |
| ★★ | As above, but structured format | Spreadsheet, Excel, PC-Axis |
| ★★★ | As above, but unpatented format | CSV, SDMX |
| ★★★★ | As above + metadata, hyperlinks | RDF/XML |
| ★★★★★ | As above + hyperlinks to other datasets creating context | RDFa with external hyperlinks |

The costs and benefits on different stages of open data development are not limited to the technical aspects. Far more difficult and questionable are the issues of social costs and benefits of open government. It should be noted that the five-star model is also applied in Polish government. It is included in the aforementioned Program of Opening of Public Data prepared by Ministry of Digital Affairs (2016a). Moreover, the requirements of Operational Program Digital Poland 2014-2020, promotes projects that have at least three-stars on the scale of "Five Star Open Data."

To fully exploit the possibilities of open data, they should be settled in the appropriate context that allows for creating new knowledge, while using this knowledge by third party services and applications. Open datasets are the essential raw material from which value could be extracted. Business models based on open data rely on the transformation of raw data sets into information, knowledge, and understanding in accordance with the idea of Ackoff's pyramid. Ackoff (1989) suggested that data in itself has little value if information, knowledge, and understanding cannot be gained from it. Therefore, publishing institutions play a critical role in the transformation process from data to knowledge. In particular, good practices are necessary, such as publishing content in open formats, in an easy to use manner and providing metadata informing about the context of the data creation.

# Exploratory Review of the Content of CRPI

## Open Government Data in Poland

The significance of open public data is recognized also by the Polish government. In the program Opening up of Public Data prepared by the Ministry of Digital Affairs (2016a) it was noted that the huge amount of data produced and gathered by public administration may be fundamental for development of innovative goods and services. Open data has the potential to stimulate economy by creating new jobs and encouraging investments in creative industry. Moreover, the access of citizens to open data and information is the basic instrument of public control over activities of government. However, the practice looks different. As the OECD has estimated in the report "Open Government Data Reviews - Poland - Unlocking the Value of Government Data" (OECD, 2015), OGD in Poland today is at a very early stage of development. Compared to other OECD countries, Poland ranks very low in effective government support for the development of OGD. The reasons for this are:

- The relatively low availability of useful content, i.e. basic datasets determined by the G8 Charter on Open Data;
- The relatively low level of accessibility of data on the national CRPI portal due to inconvenient formats, lack of good tools and functionalities;
- Little proactive government support to foster innovative reuse and stakeholder engagement in this area.

The described concept is particularly addressed by two legal regulations in Poland:

- Act of 6 September 2001 on the access to public information (Ustawa o dostępie do informacji publicznej),
- Act of 25 February 2016 on the re-use of public sector information (Ustawa o ponownym wykorzystywaniu informacji sektora publicznego).

Practice and jurisprudence emphasize different specifics of the right to public information and its reuse. The first one belongs to the category of rights to independence and allows for access to knowledge about activities of public authorities (transparency of the activities of public administration). The second one is economic law, which gives the right to create added value by using information gathered by public sector (Gałach, Kędzierska, Lipiński, Opaliński, Pietrzak, Szustakiewicz, & Zołotar, 2015).Although the idea of Open Government is strongly motivated by economic concerns, it cannot be detached from the right to public information. The first of the mentioned acts imposes publication of public information in the Bulletin of Public Information and central repository.

The Polish legal regulations mentioned above implement the Directive 2003/98/WE of the European Parliament and of the Council of 17 November 2003 on the reuse of public sector information. The Directive emphasizes the evolution towards information and knowledge society and the role of digital resources in this evolution. According to the Directive and Polish regulation "reusing" means using the documents owned by public sector by natural or legal persons with the commercial or non-commercial purpose other than originally intended (the public tasks). Public sector bodies make their documents available in preexisting format or language through electronic means where possible and appropriate (Directive 2003/98/EC).

## CRPI in Poland in Comparison to Other Countries

We decided to make comparison of data accessibility in the Polish CRPI with similar services of other countries. In Polish conditions, the idea of OGD is thought to be implemented by Ministry of Digital Affairs by the Public Data service (http://danepubliczne.gov.pl/). The aim of this service is to gather in one place the data of special importance for the development of innovativeness and information society in the country. The service is dedicated as well to public administration as to business and citizens. Its fundamental objective is the implementation of Central Repository for Public Information. In January 2018, 114 units of public administration published 867 datasets in CRPI, which were divided on nine categories. During the investigation period from March 2017 to January 2018, the number of datasets increased by 25%. At that time, we observed a change in the number of datasets made available by particular institutions

The comparison was made on the example of USA as one of the leaders in open data. This country had been a pioneer in the area of open government data. According to some rankings another leader is UK and Germany (respectively 11th & 21st position in Open Data Maturity in Europe report), as another big European country having large amount of public data similar to Poland. In the USA the service data.gov performs similar role to the Polish CRPI. Data.gov currently contains 231,138 datasets divided on 14 categories primarily provided by Federal Government (186,311 datasets). By collaborating with non-federal data sources, Data.gov is able to include this data in the catalog. For example, currently it provides 16,271 datasets from State Government, 9,479 from Local Government, 6,683 from City Government and 1,950 datasets from Country Government. It should be noted that during the 10 months of research, the website has clearly seen the largest increase in the number of datasets made available by around 70% among other surveyed websites.

The European websites surveyed provide significantly less data. British CRPI is a service (http://data.gov.uk/) where 1,414 publishers of central administration, local authorities and private sector make their data available. Currently, there is 43,620 datasets. The role of CRPI in Germany is performed by govdata.de. For now, it contains 20,650 datasets, which are divided on 13 categories. During the last ten months the amount of data has increased by less than 10% in both of the examined European services. For the research purposes nine popular formats have been selected. Next, the percentages of datasets in those formats have been calculated in relation to all examined available formats. Some of the data is available simultaneously in different formats. The comparison of popularity of particular data formats in CRPI for the surveyed countries is presented in Table 3.

**Table 3.** Popularity of data formats in CRPI of different countries (January 2018 status)

| State | Amt. of datasets | html | pdf | csv | xls<br>xlsx | xml | json | gml | rdf | wms |
|---|---|---|---|---|---|---|---|---|---|---|
| **Poland** | **867** | 9% | 16% | 21% | 34% | 9% | 6% | 3% | 0% | 1% |
| **USA** | **231 138** | 36% | 20% | 8% | 2% | 16% | 6% | 1% | 4% | 5% |
| **UK** | **43 620** | 39% | 5% | 24% | 7% | 3% | 4% | 0% | 1% | 16% |
| **Germany** | **20 650** | 26% | 10% | 26% | 14% | 5% | 1% | 1% | 0% | 16% |

## Analysis of Application of Open Data Concept Under Polish Conditions

Primary data providers are government administration bodies, national research institutes, Special Purpose Funds, and other administrative legal persons, particularly Social Insurance Fund (ZUS), National Health Fund (NFZ). In practice, important data providers are also local governments and to small extent non-governmental organizations. In March 2017, there were 90 providers and 693 datasets. It has to be noted that the number of datasets is controversial, because the notion of dataset encompasses data of different characteristics. Large databases contain many tables and high number of records, sometimes the data sources are links to external applications, often single tables of spreadsheets or text documents such as Portable Data Format (PDF) or some text editor.

The largest number of datasets is provided by Central Statistics Office (GUS) (70 datasets) among which there are statistical yearbooks containing PDF documents with spreadsheets attachments. At the same time, the datasets of GUS are accessible by web services (Application Programming Interface (API)), these are registers of public data TERYT and REGON. The Office of Technical Inspection (UDT) for example provides only two datasets, of which the first is several pages PDF document with contact data of its local branches. The second dataset of UDT is the list of domestic companies certified by UDT.

Data providers can be divided in five categories: government authorities, particularly ministries (19 providers), other central public authorities (24), research institutes (7), local governments (34), other central institutions and special purpose funds (6). Government authorities provide jointly 273 datasets, however, only in two cases the API tool is available. The largest number of datasets (37) is provided by Ministry of Justice. These datasets mainly pertain to judicial proceedings, court appearances, etc. The data format is most often comma separated value (CSV) file, which allows for classifying them as three stars. A similar number of 35 datasets is published by the Ministry of Science and Higher Education, these are different data on Universities, mostly published as Microsoft Excel (XLS) files, which can be classified as two stars. The third institution in terms of number of provided datasets (27) is the Ministry of Maritime Affairs and Inland Waterway. However, these registers are few or even single documents, which are mostly hyperlinks to other websites containing proper documents. Such situation let us to classify this group as one star. In the described group of public administration there is no dataset that could be classified higher than three stars. API is available only in case of Ministry of Digital Affairs (access to the data on regional broadband networks) and Ministry of Development (access to the Central Registry and Information about Business Activity).

In the category of other central administration institutions, the attention should be paid to Central Statistics Office because of the number of provided datasets (700) of which three are accessible through API. The second highest number of datasets is provided by the Office for Competition and Consumer Protection (UOKiK) (26), they are available mostly as Microsoft Excel files (XLS/XLSX) or CSV. The Head Office of Geodesy and Cartography (GUGiK) deserves particular attention as it provides eight datasets in the most advanced form. In case of the four of them there are APIs available. In the case of the State Register of Geographic Names, the linked data format (RDF) is used, which allows for classifying the dataset as five stars. Moreover, there are alternative methods of access to data provided. Particularly there is a hyperlink to specialized portal of spatial information – geoportal.gov.pl, which also implements the idea of open data. Regarding the methods of data publishing it should be noted that National Heritage Institute provides seven datasets of which four are accessible by API. For the two datasets – register of historical monuments and register of objects included in the UNESCO World Heritage List – the linked data format RDF is available. However, the both mentioned datasets are relatively small and rarely updated.

The research institutes usually make available only single datasets. The exception is the Institute of Agricultural and Food Economics, providing 15 datasets, however, they contain mainly PDF files and hyperlinks to the website with reports on agriculture market in PDF files. In the category of other central institutions and special purpose funds, the attention should be paid to Social Insurance Fund (ZUS) because of large number (64) of datasets provided. Most commonly these are Excel files (XLS/XLSX) and PDFs. The National Library, offers only two datasets, but both are accessible by API – it is the bibliographic database and portal polona.pl representing digitized assets of National Library and cooperating institutions. Two APIs are also provided by Foundation: Project: Poland making available the datasets of the Open Monuments.

**Table 4.** The characteristic of Polish CRPI

| Category of institution | Amt. of inst. | Amt. of datasets | Amt. of APIs | HTML | PDF | CSV | XLS XLSX | DOC DOCX | XML | JSON | GML | RDF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Government administration bodies | 19 | 273 | 2 | 34 | 59 | 69 | 132 | 8 | 0 | 1 | 0 | 0 |
| Other bodies of central administration | 24 | 229 | 12 | 19 | 61 | 48 | 94 | 7 | 17 | 0 | 3 | 3 |
| Research institutes | 7 | 23 | 0 | 18 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| Local government | 34 | 93 | 75 | 4 | 5 | 29 | 4 | 0 | 57 | 45 | 29 | 0 |
| Other institutions | 6 | 75 | 4 | 1 | 24 | 1 | 63 | 2 | 1 | 4 | 0 | 0 |
| **Total** | **90** | **693** | **93** | **76** | **151** | **149** | **295** | **17** | **75** | **50** | **32** | **3** |

Seemingly the group of local government data providers is quite large. However, 26 bodies are institutions of municipalities or small cities, which publish only contact data by the service www.punktyadresowe.pl. It is remarkable that for these datasets there is the API available. The only local governments providing more datasets are offices of the five metropolises: Warsaw, Wroclaw, Poznan, Lublin, and Gdansk. The first four of the mentioned, partially use API standard. For example, timetables of urban public transport are often published that way.

In case of all the providers (based on service data) the most popular are files in Microsoft Excel, PDF, and CVS formats. Less common are formats required for using API (XML, JSON, & GML). Very rarely applied is the RDF format typical for linked open data. Often the repository contains only the hyperlink to external website (in HTML). Detailed data with the regard to categories of bodies with the number or their datasets, APIs, and popularity of selected data formats are presented in Table 4.

A partial analysis of the open formats used in Poland has been made by Ministry of Digital Affairs (2016a) under the Program of Opening of Public Data. It has been noticed that in Poland there is a tendency to publish data in unsearchable formats, (PDF, PNG/JPG) and relatively small share of structured formats (XLS, CSV). Unsearchable formats should be used only when there is a need to present hard copy of a document or a graphical object.

## Discussion

Our research indicates the difference in the number of datasets made available is significant, in which the USA is the decisive leader, Poland remains significantly behind other analyzed countries. Comparing the available data formats, it can be seen that there are significant differences between the Polish CRPI and American, British, or German repositories. In comparison to USA, in Poland the XLS/XLSX formats are used more often and the number of linked data format (RDF) and XML, JSON typical to API is significantly less common. In Poland, the references to external websites (in HTML) are clearly less used than in other countries. It can be assumed that the above analysis proves greater maturity of open data repositories in USA than in Poland, UK, and Germany.

The quantity and format of the data available is not the only factor in assessing the maturity of data openness in OGD. Rankings attempting to evaluate the maturity of open data in different countries are ambiguous. According to Open Data Barometer report (Web Foundation, 2015) the leaders are UK and USA. The Barometer ranks nations on:

- Readiness (How prepared are governments for open data initiatives? What policies are in place?),
- Implementation (Are governments putting their commitments into practice?),
- Impact (Is open government data being used in ways that bring practical benefit?).

UK and USA occupy the first two places. Poland is on 32nd place among the 92 countries classified.

The Report Open Data Maturity in Europe 2016 (European Union, 2016) evaluates UK only at 14th place, and Poland on the 17th place among 31 surveyed countries (EU plus Norway, Switzerland, & Liechtenstein). Two fundamental factors of the evaluation are:

- Open Data Readiness (presence of open data policy, licensing norms, extent of coordination at national level, use of data, and impact of open data)
- Portal Maturity (the usability of the portal regarding the availability of functions, the overall reusability of data such as machine readability and accessibility of datasets)

Regarding the second factor, Poland takes 24th place, clearly below the average for the EU, and UK is on the 14th place. These rankings indicate a difficulty in reliable evaluation of open data. In this paper the focus has been put on popularity of particular data formats in CRPI, because it is one of the basic factors determining reuse of information.

The problems of OGD in Poland are recognized by The Ministry of Digital Affairs (2016b), therefore, the following activities are planned in the Program of Integrated Computerization of State:

- Providing access to public e-services by the Electronic Platform of Public Administration Services (ePUAP), branch and regional services
- Inventory of all the public data assets belonging to government bodies
- Providing access to public information via API
- Promoting the reuse of published open data

The goals are defined quite generally, but are closely related to CRPI. Making data available in CRPI requires prior inventory of the data. The consequence of making the data available is the possibility to reuse them and the reuse of data is facilitated by the use of API.

Our research results in this study show that the data formats characteristic for maturity of open data and APIs are rarely used. Similar conclusions flow from the reports developed by Polish authorities. A partial analysis of the open formats used in Poland was made by The Ministry of Digital Affairs (2016a) under the Program of Opening of Public Data. It had been noticed that in Poland there is a tendency to publish data in unsearchable formats, such as PDF and PNG/JPG and there is a relatively small share of structured formats XLS and CSV. Unsearchable formats should be used only when there is a need to present hard copy of a document or graphical object in digital form. To facilitate reusing it is recommended that the public data and metadata were:

- Prepared in a user-friendly way, using communicative and understandable language
- Made available in the machine-readable format and open formats to enable reuse
- Described by metadata
- Published and updated on the possibly lowest level of aggregation, the exception is when there are privacy and data protection issues
- Published and stored in a stable location
- Available also by programming interfaces (i.e. APIs)

We generally agree with the diagnosis and recommendations, however the comparison with other countries shows that the spreadsheet formats are frequently used. The fact is also that the closed data formats (XLS, XLSX) are overused in Polish open data repositories, while the open format CSV should be used instead. The main problem, however, is the small amount of open data published in Poland. Therefore, percentage comparisons do not show sufficiently Poland's delay in providing open data.

## Conclusions

The idea of open data plays a significant role for information society. These data are public property and if there are no contraindications (such as public security, privacy) they should be publicly accessible. The mentioned contraindications may, however, severely limit the data openness. Depersonalization, which should assure the right to privacy and trade secret, can be insufficient in case of big data analysis combining data from many sources.

Different aspects of data accessibility may be considered, such as:

- Control – public scrutiny of government is an important tool of democracy
- Economic – open data is a fuel for development of data processing businesses which contributes to development of the whole economy
- Personal – supporting daily life of individuals by the access to public data

The mentioned reports (Web Foundation, 2015; European Union, 2016) show the difficulty and ambiguity of evaluation of the maturity of open data in different countries. Comparing the datasets is difficult due to the very different amount of data contained in them. There is no point in comparing the volume of shared data due to different formats of data sharing. Discrepancies between reports can be caused by different focus of the evaluation, for example on the way of publishing or on the possibility of data reuse.

The statistics reports and own research results described in this paper show different barriers impeding implementation of the open government in Poland. These barriers are:

- Unwillingness to make public data accessible, which is related to lack of awareness of the importance of such data and the attempt to reduce the public scrutiny of administration of different levels
- Legal environment of open data and existence of legislation on public data access and the rights to reusing public information. The level of detail of legal acts is very important as well as their accurateness and technical context
- Lack of technical solutions facilitating taking advantage of open data, such as data repositories, proper formats and programming interfaces

The research that we conducted shows the very different approaches of Polish government and public administration bodies to making their data available. It can be assumed that the legal regulations regarding open data are inconclusive or lack proper standards. It is easier to evaluate the issue of publishing data, which is regulated by law than the possibility of reusing data. At the same time, it can be seen that there is awareness of the problem and significance of public open data in government bodies. The participant observation in the projects of computerization of public administration and the analysis of strategic documents show that the idea of open data is successfully implemented by some of the local governments.

The research presented in this paper and other cited sources show that Poland is still lagging behind the leaders of opening public sector data. The basic problem in Poland is the small amount of data available in CRPI. At the same time currently, the issue is application of open and useful data formats (now mainly unstructured & patented formats are used) and making available more programming interfaces to access the data by third party applications. The technical base for API of danepubliczne.gov.pl catalog is the Comprehensive Knowledge Archive Network (CKAN), a web-based powerful open source data platform. CKAN is used worldwide and powers a variety of official and community data portals. The next step in the development of open data is the practical implementation of linked data in OGD repositories. Recently, linked data–based solutions have been adopted by the leading practitioners (such as Data.gov in the US & Data.gov.uk in the UK).

At the current stage of research, we are not yet able to unambiguously answer the question about the practical usefulness of the published data and the actual scale of their use. For this purpose, the projects observed must be finished. Further observation will, in particular, include the development of IT tools using OGDs, connecting them with each other and re-sharing such combined data. In future research we intend to follow the changes in the quantity and method of OGD provisioning in Poland and other countries. Emphasis will be put on the practices of linked data sharing and reuse.

## References

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applies Systems Analysis, 16*, 3-9.

Act of 6 September 2001 on access to public information.

Act of 25 February 2016 on the re-use of public sector information.

Banisar, D. (2005). *Effective open government: Improving public access to government information*, Paris: OECD Publishing.

Berners-Lee, T. (2006). Linked data. Retrieved from https://www.w3.org/DesignIssues/LinkedData.html

Bizer, C., Heath, T., & Berners-Lee T. (2009). Linked data – The story so far. *International Journal on Semantic Web and Information Systems, 5*(3), 1-22.

Constitution of the Republic of Poland of 2nd April, 1997.

Curristine, T., & Abbott, B. (ed.) (2005). *Modernising government: The way forward*. Paris, France: OECD Publishing.

DeWalt, K.M. & DeWalt, B.R. (2011). *Participant observation: A guide for fieldworkers* (2nd ed.). Lanham: AltaMira Press

Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the reuse of public sector information.

European Commission (2016). *Open data maturity in Europe*. Retrieved from: https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n2_2016.pdf

European Union (2016). *Open data maturity in Europe 2016. Insight into the European state of play*. European Union. Retrieved from: https://www.capgemini.com/consulting/resources/open-data/

Gałach, A., Kędzierska, K., Lipiński, A., Opaliński, B., Pietrzak, B., Szustakiewicz, P., & Zołotar. A. (2015). *Dostęp do informacji publicznej a prawo do prywatności*, Warszawa: Wydawnictwo C.H.Beck.

Hamilton, G., & Saunderson, F. (2017). *Open licensing for cultural heritage*. London, UK: Facet Publishing.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, *29*(4), 258-268.

Kawulich, B. (2005). Participant observation as a data collection method. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 6(2), Art. 43, Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/466

Kitsios, F., Papachristos, N., & Kamariotou, M. (2017). *Business models for open data ecosystem: Challenges and motivations for entrepreneurship and innovation*, *Proceedings of 19th IEEE international conference on business informatics (CBI'17)*, Thessaloniki: IEEE, 398-408.

Kozierski, P., Kabaciński, R., Lis, M., & Kaczmarek, P. (2013). *Open access. Analiza zjawiska z punktu widzenia polskiego naukowca*, Poznań, Kraków: Impuls.

Krishnaswamy, K. N., Sivakumar, A. I., & Mathirajan, M. (2009). *Management research methodology: Integration of principles, methods and techniques*. New Delhi: Prentice Hall.

Lakomaa, E., & Kallberg, J. (2013). Open data as a foundation for innovation: The enabling effect of free public sector information for entrepreneurs. *IEEE Access*, *7*, 558-563.

Malinowska-Misiąg, E. (ed.) (2016). *Jawność i przejrzystość finansów publicznych*, Warszawa: Oficyna Wydawnicza Szkoła Główna Handlowa w Warszawie.

Mayer-Schonberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. London, UK: John Murray.

Meijer, A. (2009). Understanding modern transparency, *International Review of Administrative Sciences*, *75*(2), 255-269.

Mendel, T. (2003). *Freedom of information: A comparative legal survey*. New Delhi: UNESCO, p. 79.

Ministry of Digital Affairs (2016a). *Program otwierania danych publicznych*, Retrieved from https://www.gov.pl/documents/31305/0/program_otwierania_danych_publicznych.pdf

Ministry of Digital Affairs (2016b). *Program zintegrowanej informatyzacji państwa*, , Retrieved from https://www.gov.pl/documents/31305/0/program_zintegrowanej_informatyzacji_panstwa_1.pdf

OECD (2015). *OECD Open government data reviews. Poland. Unlocking the Value. of Government Data. Assessment and proposals for action.* Paris: OECD, Retrieved from http://www.oecd.org/gov/Open-Government-Data-Review-of-Poland-Assessment-and-Recommendations.pdf

Orszag, P. (2009). *Open government directive: Memorandum for the heads of executive departments and agencies*. Retrieved from https://obamawhitehouse.archives.gov/open/documents/open-government-directive

Papińska-Kacperek, J., & Polańska, K. (2015). Analiza zaawansowania realizacji idei Open Government Data w wybranych krajach, *Zeszyty Naukowe Uniwersytetu Szczecińskiego Nr 874, Studia Informatica", 37*, 103-114.

Pollers, A., Amato, C. D., Arenas, M., & Handschuh, S. (ed.) (2011). *Reasoning web. Semantic technologies for the web of data – 7th International Summer School 2011*, Galway, Ireland: Springer, p. 29.

Stagars, M. (2016). *Open data in southeast Asia. Towards economic prosperity, government transparency and citizen participation in the ASEAN*. Singapore: Palgrave Macmillan.

Szupiluk, R. (2013). *Dekompozycje wielowymiarowe w agregacji predykcyjnych modeli Data Mining.* Warszawa: Oficyna Wydawnicza SGH.

W3C (2014). *Best Practices for publishing linked data W3C working group note*. Retrieved from https://www.w3.org/TR/ld-bp/

Wood, D. (ed.) (2011). *Linking government data*. New York–Dordrecht–Heidelberg–London: Springer, p. 138.

Web Foundation (2015). *Open data barometer. ODB Global Report. Third Edition*, Web Foundation. Retrieved from http://opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-GlobalReport.pdf

https://cppc.gov.pl/

https://danepubliczne.gov.pl/

https://data.gov.uk/

https://index.okfn.org/

https://opengovernmentdata.org/

https://www.govdata.de/

https://www.data.gov/

http://thegovlab.org/govlab-index-on-open-data-2016-edition/

https://theodi.org/knowledge-opinion/reports/

https://www.digital-science.com/resources/portfolio-reports/state-open-data-2017/

https://www.elsevier.com/about/open-science/research-data/open-data-report/

https://www.oecd-ilibrary.org/governance/government-at-a-glance-2017/open-useful-reusable-government-data-index-ourdata-2017_gov_glance-2017-graph139-en/

https://www.opengovpartnership.org/

https://www.netmarketshare.com/

https://www.tomforth.co.uk/globalopendata/

https://www.transparency.org/news/feature/open_data_promise_but_not_enough_progress_from_g20_countries/

## Authors' Biographies

**Jędrzej Wieczorkowski** is an assistant professor in the Institute of Information Systems and Digital Economy at the Warsaw School of Economics. He is also an independent IT project consultant and an expert evaluating such projects. His research interests include big data and business intelligence applications and the consequences of using these methods, in particular behavior of IT users and privacy problem.

**Ilona Pawełoszek** is an assistant professor at the Faculty of Management of Częstochowa University of Technology. Her professional interests include applications of Semantic Web and Linked Data in business and public administration.