Open Government Data

Small Country User's Perspective

Mladen Varga, Katarina Ćurko Faculty of Economics & Business University of Zagreb Zagreb, Croatia mvarga@efzg.hr, kcurko@efzg.hr

Abstract -The paper describes how users in a small transitional country perceive open government data. It discusses the objectives of opening government data to the public, considers what end-users require, and investigates impediments to enduser's adoption of open data. The results reveal the following: end-users are generally unprepared to use open data; they have no confidence in the credibility of published data; end users do not differentiate between open data and private data; they are not familiar with the purpose of open data and opportunities arising from using them. Finally, the paper discusses possible government measures to boost open data usage. The following measures are proposed: a single open data portal should be organized; data suppliers should be stimulated to publish, certify and maintain their open data; and end-users should be encouraged to use the data.

Keywords-open data; open government data

I. INTRODUCTION

The concept of open data has existed for some time. It corresponds to open movement, such as open access to scientific information, open source software etc. In this paper, we shall mainly consider government data (GD), also referred to as public sector information (PSI). Government bodies collect, produce, reproduce and disseminate public data in many areas of activity while accomplishing their institutional tasks. These data, such as geographical information, statistics, weather data, transportation data, public health data, etc. are of public interest, belong to the whole community and every citizen is entitled to know and use them. Government data could be opened for re-use by citizens, business and industry, science, media, civil society and others. The benefits of opening government data are numerous, even it is impossible to predict what value it will create in the future [1].

To achieve their full potential, government data must be open. Making data open could enormously increase civil sector engagement. An effective public information system should include a considerable amount of open government data and be equipped with efficient tools for data analysis and visualization.

The paper describes the position and the early experience of an open data initiative in a small transitional country (the Republic of Croatia with about 4 million inhabitants). It discusses the objectives of opening government data to the public, considers what end-users require, investigates Tomislav Vračić Ministry of Public Administration Zagreb, Croatia tvracic@uprava.hr

impediments to end-user's adoption of open data, and possible government measures to boost open data usage. It is organized as follows. The definition and types of open data are discussed in Section II, end-users' requirements as seen by the average user are described in Section III, the main impediments to citizens' adoption of open government data are considered in Section IV, and possible government measures for boosting the usage of open data are described in Section V and summarized in Section VI. The findings listed correspond to the specific situation of a smaller community, in the grip of recession and not too enthusiastic about opening government data.

II. DEFINITION OF OPEN DATA

Formal definitions of open data are relatively new. Open Data Institute [2] defines open data as information that is available for anyone to use, for any purpose, at no cost. OpenDefinition.org [3] defines open data as data that can be freely used, reused and redistributed by anyone.

Private data, i.e., sensitive personal data, cannot be the subject of open data. There is no consensus on what constitutes private data, and the debate on the subject continues.

Two most prominent types of open data are open science data and open government data.

Open science data (OSD) are freely available, allowing any user to read, download, copy, distribute, print, search, index, use by software or use for any other legal purpose. When the data are reproduced or distributed the copyright must be acknowledged, author's control over the integrity of their work must be preserved and the work must be cited. Public availability and reusability of scientific data leads to transparency, fairness and increased quality of science. Open science data are creating value in science itself.

The same is true for open government data (OGD), which we are considering in the paper. Open government data are data produced by government or its bodies, that can be freely used, reused and redistributed by anyone, resulting in a fairer and better government. As discussed in the literature [1], [11], [13], reasons for making data open include: (a) transparency and accountability, (b) innovation and economic development; and (c) inclusion, empowerment and law enforcement.

A. Transparency and Accountability

Open data increase transparency. Free access to government data can empower citizens to exercise their democratic rights [11]. Consequently, transparency increases the accountability of government and its bodies. It is reasonable to expect that they will make better decisions in the public interest. The focus is on the political domain. Open data create value in government itself by increasing its efficiency.

B. Innovation and Economic Development

Open data may enable innovators to improve services or build new products and services within public or private sector. Open data may shift certain decision making from the state into the market [1]. The key focus is on the economic domain. Open data create value in many ways, for example by helping create new products or services in places where they are missing, including the development of new products built directly on PSI [13].

A type of open data, which may be considered as OGD in a very broad of sense, is open business data (OBD). The key focus is also on the economic domain. The source of data is the business sector. Examples include data collected at the chambers of commerce or trade, statistical or marketing agencies or in corporations that are willing to open their data to general public.

C. Inclusion, Empowerment and Law Enforcement

The motivation to open government data is to involve citizens in policing and law enforcement [11]. Open data may remove power imbalances that resulted from asymmetric information, bring new actors in the political debate, especially those with special interests or needs. The key focus is on the social and law domain.

III. USER REQUIREMENTS ON OPEN GOVERNMENT DATA

Who are the typical stakeholders and users of open data? They are not a uniform group of people who share the same urge and interest to access data. In [5] they are classified as public sector, private sector, donors, civil society organizations, academia, civil hackers, and media, depending on their role in open data initiatives. Based on data driven classification [5] they are classified as data producers, data consumers, data intermediaries and data specialists. The Table I presents the different types of stakeholders, based on their expected main roles and tasks.

What do users, i.e., data consumers, expect from open data? They want to have access to (a) all types of useful data, (b) in a single place, (c) that the data are easily accessible, (d) appropriately described and interpreted, and (e) free or almost free. The findings are based on a number of unstructured interviews with potential users, and public consultations with non-governmental organizations, business sector and academia, both of which were conducted in the preparatory phase of the Croatian government's e-Citizen project [23]. Then established, the Open Government Partnership helped to define the expectations from open data.

A. All Types of Useful Data

Users will access the data they need for various purposes. Each users' community is unique in its own way and has its own needs and priorities in accessing data. We are discussing data consumers' needs and priorities in Croatia. Although a member state of the EU from July 2013, Croatia is in many aspects a transitional state. The needs and priorities of data consumers in transitional states may be different from the ones in highly democratic EU member states.

The Croatian Parliament recently adopted the new Right to Information Act [15], aligned with the *PSI Directive* [16], and a process of alignment with the *PSI Directive revision* [17] is also planned to be accomplished in next two years. The Open Government Partnership Action Plan for the implementation of the Open Government Partnership initiative in Croatia 2012-2013 [24] has also been created. Consequently, according to the Right to Information Act and other regulation acts, e.g. Public Procurement Act, State Budget Act, etc., a number of datasets has been made available to the public. Open government data web portal is not yet available, although it is being developed [25].

	Data producers	Data consumers	Data intermediaries	Data specialists
Public sector	Government data producing	Government interoperability		Topic specialization (health, transportation, education, etc.)
Private sector	Business data producing	Data market	Data consultancy	Cross-topic specialization (legal, economic, entrepreneurship, etc.)
Donors, Foundations and International Organizations	Social data producing			
Civil Society Organizations	Mainly social data producing	Data for social good	Data advocacy	
Academia and Research	Science data producing	Data analytics	Data research	Data science
Civic Hackers		Civic data apps development	Apps and visualizations development	
Media		Data journalism		

TABLE I. STAKEHOLDERS' CLASSIFICATION

Although citizens are increasingly aware that greater government transparency and accountability are necessary, government is still not transparent enough. According to the results of online interviews with individuals working in public administration in Croatia, most of them are willing to provide open data to the public.

The state suffers a long recession. The society is encumbered by corruption (Croatia is ranked 62nd according to the Transparency International Corruption Perceptions Index [18]). Opening data to the general public is therefore crucial and of highest priority.

All reasons to make data open apply as we need to achieve (a) transparency and accountability, (b) innovation and economic development; and (c) inclusion, empowerment and law enforcement. Most Croatian citizens list these goals in the same order.

B. In a Single Place

Users will prefer to access open data through a single neutral place, i.e., a web portal, where they can find all the available open data. The neutral place, web portal, may be owned by government but it must not mirror any kind of politics, government bodies' relationships, election cycles and other obstacles which will prevent users from trusting open government data.

Who owns and maintains data is a very important question. The owner or steward of open data could be either a web portal or a public (government) agency. Stewardship of open data is particularly important as it manages the quality and timeliness of open data and related metadata.

Following legal requirements, the Croatian Parliament recently set up the position of Information Commissioner [19] whose expected involvement in the implementation of open data will raise confidence, at least in the case of Croatia.

The Croatian Government additionally investigated the treatment of public data by public authorities and was able to prove that it is essential to develop a single open data portal. Since a single public administration portal is being built under the domain gov.hr as a one-stop-shop for all users, it would be logical to have open data published under the same domain, for example, data.gov.hr. This would be in line with similar naming practices for national portals in the EU and the rest of the World (e.g. data.gov.uk, data.gov.it, data.gv.at, data.gov, etc.).

A useful model for using and managing a web portal is shown in Fig. 1 [6]. The model consists of the data user finding data, the data portal acting as single point of access



Figure 1. Role of data user, data portal and data supplier

to open data; and the data supplier producing and collecting data. The left-hand side of the Fig. 1 shows the model of direct data provision. The data supplier produces data and collects them (1) at the data portal. The data are published on the data portal (2) and located at the data portal. Users find (3) and obtain (4) data from the data portal. The right-hand side of the Fig. 1 shows the model of indirect data provision. The data supplier produces, collects and publishes data (1). The data are located at the data portal (2). At the data portal, users find (3) data obtained (4) by the data supplier.

In both models, metadata are located at the data portal making data retrievable through the data portal. In direct data provision model the data are settled at the data portal, and in indirect model at the data supplier. Both models can be implemented on the same data portal, leaving data suppliers to choose between the direct and indirect model of collecting data or metadata.

After starting open data initiative the majority of the open data will be governmental data. There are a lot of data outside governmental sources that may be opened. For instance, business data collected at the chambers of commerce or trade could be opened. In a small community, such as a city, county or even a small country like Croatia, it may be appropriate, at least for budgetary reasons, to have a single place to access all open data.

C. Easily Accessible and Linkable to Other Data

In addition to access through a single web portal, the ease of access is of crucial importance. Open data are easily accessible if users can approach all open data through a central metadata portal. It is essential that the metadata portal represents the full repertoire of open data consistently and clearly. A good proposal for metadata repertoire and classification of themes covered by open data is described in [5].

Most of current OGD portals make data available to users of the web portals as downloadable files in formats such as pdf, xls, csv, xml, json etc. Making data available as linked data (LD) through RDF model and RESTful APIs or SPARQL search interfaces is not so common although linked data offers the best practices for publishing and linking data on the web.

To benefit from open data, it is important to allow linking disparate open data sets. The linked open data (LOD) can serve as a platform for new knowledge and can enable new services or applications. Services and application are easier to develop in a LOD environment. Tim Berners-Lee, initiator of linked data, suggests a five stars classification scheme for open data:

- one star: Information is available on the web in any format under an open license
- two stars: Information is available as structured data (e.g. in Excel instead of an image scan of a table)
- three stars: Information is available in nonproprietary formats (e.g. csv instead of Excel)
- four stars: URI identification is used so users can point at individual data

• five stars: Data are linked to other data to provide context. Network effect is achieved.

Achieving full benefit of linked open data can be obtained in the following steps: (a) Analyzing and cleaning data, modeling them by choosing established vocabularies, conversion data to RDF [7], (b) creating an unifier resource identifier (URI) for each data object, (c) choosing appropriate vocabularies for open data or creation of own vocabulary, (d) specification of licenses for re-using data, (e) conversion to RDF. Before publishing, data may be linked to other data, which may increase their value.

The three stars level is considered the minimum for the release of government data for re-use: non-proprietary, machine readable, and accessible via the web. Benefits of linked open data, which are data with four and five stars, include possibilities of linkage to them from any other place, possibilities of bookmarking them and re-using parts of them.

D. Appropriately Described

Data should be thoroughly described in order to explain the problem or area to which they relate. The open data description includes metadata described in user's words explaining objects, attributes and relationships between data. The purpose of data initially collected should be explained, i.e., why the data are collected, the context of the data collection, possible applications of data, for what purpose data should not or cannot be used, etc. The data should be declared not to violate privacy principles. Raw data should include full details explaining what the data relates to, how they were collected, who collected them, and how they are formatted.

Data consumer should be aware that re-use of open data should acknowledge the source of data.

E. Free or Almost Free

Access to open data should be free. Price of Internet access should be acceptable to all classes of users.

IV. IMPEDIMENTS TO CITIZENS'ADOPTION OF OPEN GOVERNMENT DATA

The impediments to adoption of open government data are many, from political and social to technical. The paper presents a number of impediments to adoption of open government data. The evidence was collected by observing the process of opening the data in Croatia [15], [23], [24], [25] and from discussions with potential open data consumers. Our findings were compared to experiences accompanying open data initiatives in the EU, predominantly in the UK [10]. The majority of impediments noted are quite the same as in communities that have overcome the initial steps in opening data. The presented impediments are listed in the natural order, from those conceptual to technical. Here are, in our opinion, the most significant impediments:

- (1) Some citizens do not know or they do not feel that they have the right to seek public information from government and its bodies.
- (2) Many citizens are not familiar with the concept and importance of open data. The same findings are seen

in advanced communities, such as in the UK where Open data dialogue report [10] finds that participants of the dialogue "found the concept of open data to be abstract and relatively hard to engage with".

- (3) Most citizens cannot precisely differentiate between public and private data.
- (4) Many citizens do not feel that seeking public information will make any difference in practical problem solving in their community.
- (5) In some, less developed communities, citizens do not have complete confidence in the published public data.
- (6) Many citizens do not express an interest in personally exploring datasets. Rather they are more interested in the results and implications of researching open data. The principal benefits of open data are seen to be for researchers rather than the public [10].
- (7) Many citizens lack computer knowledge to access open data, and have difficulties in understanding forms and formats of published data.
- (8) Many citizens experience the web portal of open government data as not user friendly enough, but only appropriate for computer specialists who are able to work with csv, json, ogd, txt, xls and other formats.

The conceptual impediments (1-5) show that citizens need to become more strongly aware that they have the right to use public data, to enhance their data understanding and their belief in the data. Technical impediments (6-8) show that citizens need to acquire technical skills to access and use the data, or that data should be displayed in a technically acceptable manner.

V. GOVERNMENT MEASURES FOR BOOSTING THE USAGE OF OPEN DATA

To overcome any of previously mentioned impediments and boost the usage of open government data, one or more measures can be implemented. The proposed measures have been derived from experiences accompanying open data initiatives abroad [8], [9], [11], [12] and in Croatia (e.g. e-Citizen project [23], which represents a continuation and the improvement of the public reform process started in Croatia in the last decade; and Open Government Partnership [21], [24], which additionally helps to streamline priorities of opening public processes). The measures are listed in the natural order, as the measures on the conceptual level (1-11) and the measures on the technical level (12-15):

- (1) Governments often act as secrecy keepers, not openness leaders. In all countries studied [11], the closed government culture is detected as the barrier to opening data. Government should define a strategy to change its culture by looking at the experiences of leading governments, following various global, regional, citizens' or market initiatives, implementing EU directives or monitoring activities on open data.
- (2) Government should provide a useful definition and explanation of both public and open data that is

accepted by law, administration and citizens [8]. A kind of consensus is needed.

- (3) In all countries the tensions between open data policy and the privacy of citizens are recognized [11]. Therefore, the difference between open and private data must be clearly defined and explained.
- (4) In less developed communities the general public should be educated that it has the right to seek public data. By whom? Self-education, appropriate civil society activities, and clever government initiatives should change the situation. Imposing proper licensing may help [8] in understanding the right to public information.
- (5) Marketing of open data is poor. Citizens and interested parties are not sufficiently informed on government's activities on opening data. The civil society and government itself should do their best to market open data; explain pros of using and re-using open data.
- (6) In addition to standard marketing of open data, government should educate citizens to understand and use data [8]. Sharing useful stories and cases of successful open data usage would be helpful. Similar goals have been set in the Open Data Support project [20]. It should open a channel for communicating users' ideas, proposals for improvement of data or publishing new data, useful successful or unsuccessful cases, impediments etc. through social media or forums. Publisher of open data must be prepared to respond to users' communication.
- (7) Users should be attracted by selecting high value datasets. Datasets may be considered to be of high value if they satisfy the following criteria: data publication is mandatory by law (e.g. regulated by acts, directives, tenders or budget data); the data result from a primary governmental activity (such as health, transportation, financial data); the high level of preparedness of data (already online, such as weather data), or data of high value in general (such as business data). The focus should be on local, specific issues to raise interest for open data, at least at the moment of introducing open data [8].
- (8) Non-governmental organizations, charity organizations and business associations have to be involved in open data initiatives [8], [21].
- (9) Governments should act like product developers and measure the outcome of their activity. Open data needs to be a product that will improve transparency, accountability, economic, social and other aspects of community.
- (10) Government should clearly define and explain to users the difference between public data and private data. Personal data should be confidential, and must not be violated in any way, such as by combination of related datasets. Misunderstanding open data as an attack on privacy prevents full adoption of open data. Government should define and explain to users what open data as opposed to private data are.

- (11) In communities where citizens do not have full confidence in the published public information, data openness should be governed through an independent body.
- (12) The quality of open data is often questionable or below acceptable limits to permit the publication of data. Even in countries with experience in open data initiatives it is reported [9] that "information is often treated as a black box in the open data movement, information is often seen as a given, used uncritically, and trusted without examination, open data was collected or created for other purposes, it substantial risks for validity, relevance, and trust." When publishing open data, government has to ensure their quality. Data should be checked for inaccuracies before being opened.
- (13) Since many citizens and other users are not keen on personally exploring datasets the web portal should be equipped with as many data applications as realistically possible. A proper user friendly interface would be a catalyst of using open data.
- (14) The open data territory in not yet standardized to an acceptable level. The open data portal should use the accepted standards, and reduce number of used formats to the smallest number possible. Its graphical interface should be user friendly. The classification of themes covered by the open data portal should be adjusted to the interests of data consumers and dependent on local circumstances.
- (15) The quality of published data may vary from dataset to dataset. New practice of datasets certification such as by Open Data Institute [12] may help "data users to understand its quality, licensing, structure, and its usability, publishers understand how they can better connect with their users, etc." Certification may be localized, which is better than no certification at all.

VI. SUGGESTED ACTIONS

A general strategy of opening government data should be concentrated on actions in the following areas:

- Organizing a single open data web portal (a) supported by adequate technical and organizational infrastructure, (b) by monitoring activities on the portal and constant analysis of data usage in order to be constantly improved,
- (2) Stimulating data owners or data suppliers to publish, certify and regularly maintain their data; and
- (3) Encouraging end-users to use open data (a) considering users' feedback using forums or other appropriate tools, and (b) stimulating end-users or data suppliers to develop applications on open data and publish results based on processed data.

All of the actions above represent a challenge. They need to be well planned, performed, monitored, analyzed and continually improved.

VII. CONCLUSION AND FUTURE WORK

The key dimensions of implementing open government data as outlined by OECD [14] include challenges related to policy, technology, financing, organization, culture, and legal frameworks. Addressing these challenges is essential in a small transitional community, e.g., Croatia, if we are to create an ecosystem, and build sustainable business models for OGD initiatives that can generate the desired benefits. If not properly tackled, these challenges might obstruct or restrict the capture of benefits of national efforts aimed at spurring OGD.

Additionally, participation of Croatia in the Open Government Partnership has already helped nongovernmental organizations, charity organizations and business associations to be involved in open data initiatives.

On the other hand, there are many successful examples of countries which succeeded in opening data. Learning from these examples could streamline Croatia to faster achievement of better results in this area. It has been proposed that additional support be given to the already established collaboration of all interested parties and that already proven technical solutions be reused [22] in order to start with the open data portal in a short period of time.

The paper presented the position and early experience of an open data initiative as seen by end-users in a small transitional country. The future work may consider monitoring the development of the users' perspective on open government data, discuss development and implementation of measures for boosting the usage of open data, analyze the results of their implementation, and compare the results with solutions in other countries or communities.

References

- [1] Open Knowledge Foundation, "Open Data Handbook Documentation, release 1.0.0", 2012.
- [2] Open Data Institute, "Guides". http://theodi.org/guides/what-open-data 6.10.2013
- [3] The open definition. http://opendefinition.org/> 6.10.2013
- T.Davies, F. Perini, and J. Alonso, "Researching the emerging impacts of open data," ODDC working paper #1, 2013.
 http://www.opendataresearch.org/sites/default/files/posts/Researching%20the%20emerging%20impacts%20of%20open%20data.pdf> 25.10.2013
- [5] J. Alonso and C. Iglesias, "Open Data Directory; Use Cases and Requirements," World Wide Web Foundation, 2013. http://public.webfoundation.org/2013/06/ODD-UCR-Final.pdf> 25.10.2013
- [6] E. Kalampokis, E. Tambouris, and K. Tarabanis, "A Classification Scheme for Open Government Data: Towards Linking Decentralised Data," Int. Journal of Web Engineering and Technology, vol. 6, issue 3, June 2011, pp. 266-285.
- [7] F. Bauer and M. Kaltenboek, "Linked Open Data: The Essentials," 2012 .<http://www.semantic-web.at/LOD-TheEssentials.pdf> 25.10.2013
- [8] M. Fioretti, "Open Data: Emerging trends, issues and best practices," 2011. http://www.lem.sssup.it/WPLem/odos/ odos_report_2.pdf> 27.10.2013

- S. Dawes, "A Realistic Look at Open Data," Using Open Data Workshop Brussels, June 19-20, 2012. http://www.w3.org/2012/06/pmod/pmod2012_submission_38.pdf 27.10.2013
- [10] "Open data dialogue: Final Report," TNS 2012. <http://www.rcuk.ac.uk/documents/documents/TNSBMRBR CUKOpendatareport.pdf> 27.10.2013
- [11] N. Huijboom and T. Van den Broek, "Open data: an international comparison of strategies," European Journal of ePractice, N° 12, March/April 2011. < http://www.epractice. eu/files/European%20Journal%20epractice%20Volume%201 2_1.pdf> 27.10.2013
- [12] Open Data Institute. Open Data certificate. 27.10.2013">https://certificates.theodi.org/>27.10.2013
- [13] G. Vickery, "Review of recent studies on PSI re-use and related market developments," Information Economics Paris, <http://ec.europa.eu/information_society/newsroom/cf//docu ment.cfm?doc_id=1093> 28.10.2013
- [14] B. Ubaldi, "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives," OECD Working Papers on Public Governance, No. 22, OECD Publishing, 2013. 28.10.2013">http://dx.doi.org/10.1787/5k46bj4f03s7en>28.10.2013
- [15] "Right to Information Act," Official Gazette, No. 25/2013. in Croatian only. http://narodne-novine.nn.hr/clanci/sluzbeni/2013_02_25_403.html> 28.10.2013
- [16] "Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information on the re-use of public sector information," Official Journal of the European Union L 345/9031.12.2003 <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ: L:2003:345:0090:0096:EN:PDF> 29.10.2013
- [17] "Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information," Official Journal of the European Union L 175/1 27.6.2013 http://eur-lex.europa.eu/LexUriServ.do?uri=OJ:L:2013:175:00 01:0008:EN:PDF> 29.10.2013
- [18] Transparency International, "Corruption Perceptions Index." http://cpi.transparency.org/cpi2012/results 29.10.2013
- [19] "Decision on the election of the Information Commissioner," Official Gazette, No. 131/2013. in Croatian only. http://narodne-novine.nn.hr/clanci/sluzbeni/2013_10_131_2855.htm > 30.10.2013
- [20] Open Data Support, DG CONNECT European Commission <https://joinup.ec.europa.eu/community/ods/description> 30.10.2013
- [21] Open Government Partnership http://www.opengovpartnership.org/ 30.10.2013
- [22] Data.gov.uk To Go http://data.gov.uk/blog/datagovuk-to-go> 30.10.2013
- [23] "Decision on starting e-Citizen project" Official Gazette, No. 52/2013. in Croatian only. http://narodne-novine.nn.hr/clanci/sluzbeni/2013_05_52_1058.html> 6.1.2014
- [24] "Action Plan for the implementation of the Open Government Partnership initiative in Croatia 2012-2013," in Croatian only. http://www.uzuvrh.hr/userfiles/file/Akcijski%20plan-Partnerstvo%20za%20otvorenu%20vlast-5_4_2012_.pdf 6.1.2014
- [25] "Guidelines for making internet site," Croatian Government, In Croatian only. http://www.vlada.hr/hr/content/download/ 260703/3840158/version/1/file/Smjernice_za_izradu_internet sjedista.pdf> 8.1.2014