

Open issues of open data in Croatia

Dalibor Bužić, Jurica Đurić

Visoka škola za informacijske tehnologije, Klaićeva 7, Zagreb, Croatia

dalibor.buzic@vsite.hr, jurica.duric@vsite.hr

Optimism and euphoria about open government data as a silver bullet which should lead economy and society to significantly higher levels have deflated after few years in collision with a reality. Expectations from open data usage were great, but in exploitation it is more than questionable at which rate open data generate new economic values. Many issues appear on both sides – government as a supplier and on demand side which lags behind the supply. In this paper we analyze and comment open data published by some of the public authority bodies in Croatia and we present an overview of mobile application based on those data.

I. INTRODUCTION

Governments all over the world generate large amounts of data. Movement for free access to government data is based on the fact those data are collected and processed by taxpayers' money. The goal of free access to data is enabling citizens, companies, media and academic community to use those data for economic and social progress. Published data should be in genuine form (without aggregation), up-to-date, ready for easy machine processing, available to everyone and license-free.

Open data promoters claim that the useful value of open government data can be created by citizens, associations, companies, academic community, and government itself. Citizens can find their interest in information which helps while making decisions such as choosing city for living with regard to air pollution or climate elements (areas with high average humidity, for instance) or choosing schools for children (based on the number of children per one teacher or a success in the later stages of education) and so on. Non-government associations can improve status of vulnerable civil groups. Academic community can conduct new researches. Companies can create new services and products based on open data. At the end, whole state can have benefits – new employments in the private sector, tax incomes, raise of the economy, and one entirely different spectrum of benefits such as increasing transparency or involving citizens in activities of wider social interest.

Economic gain of open government data usage one European commission study estimates on about 40 billion euros per year in the EU [1]. But this business activity domain is relatively new, and there is a lack of knowledge about potential economic effects of usage and creating new products and services based on open data. Some of the greatest challenges for existing and new companies interested in creating new values from open data are which precisely new products and services should they make, and how to make it lucrative. Applications related to meteorological data (such as temperature, humidity, wind speed and direction, UV index) and to traffic data (such as railway timetables, delays, road works and traffic

jams) are well accepted by a wide range of users. Statistic data processed and presented in an attractive way (especially graphics) can be a good opportunity. Information can be related to pupils' success in classes and competitions, water and air quality, police crime maps or government management (expenses share by type, supplier list, miscellaneous comparisons of cities or municipalities).

II. OPEN ISSUES

Optimism and euphoria about open government data as a silver bullet which will raise the economy and society on a significantly higher level had been deflated a few years after colliding with a reality. Web site data.gov, central place for publishing open US government data, was established in May 2009. when USA (and many other countries worldwide) experienced big crisis. Expectations from open data were large, but in practice it is very questionable at which rate open data generate new economic values.

The gap between promises and barriers are formulated in five myths [2]:

- publishing data will automatically yield benefits,
- all information should be published without any restriction,
- it is enough to publish data,
- everyone can make use of open data,
- open data will result to open government.

A. Supply-side issues

Issues arise at a very first part of the value chain: government. Bad design, flawed execution and adverse consequences are three main problem areas [3]. Bad design manifests in the initial uncertainty of a concept (what do open data really mean?), unrealistic goal of absolute transparency and focus on a technology as a transparency indicator.

Flawed execution reflects in three elements: resistance in agencies, data decontextualization and expense side. Resistance appeared, among other reasons, because of unrewarding extra work in preparation of data for publishing, but also because trading data were the significant income source. The question to be answered can be, for instance, what consequences brings public access to Croatian Meteorological and hydrological service (local acronym: DHMZ) which makes earnings from selling collected data? Decreasing income can lead to cutting down investment in equipment and decreased experts' salaries. That could result in negative employee selection and, finally, in poorer service. DHMZ's income

can be compensated from government budgets, but in that case, all taxpayers suffer that load. The serious question is then: who has benefits from free data? Right now some of DHMZ's data purchasers are largest media, telecom and insurance companies. On the other side, research about opening hydrometeorologic data in Netherland revealed that at the time when data has been sold, there were ten purchasers which brought income of four million euros. After data opening, number of users has increased, and government's tax revenue was about 30 million euros, which is 750% more [4]. In any case, cost-benefit analysis should be done before opening data which generate income to government agencies.

Decontextualization is the process of converting useful complete records into useless datasets by separating data from context. Agencies in the USA, just to fulfill the 'norm', often repacked data already published somewhere else. Data were sometimes without description, and in some cases even could not be opened.

Expenses side, as the third aspect of flawed execution, is conscious or unintentional ignoring of opening data costs. Some of the costs are related to adjusting data to new standards, employee training, improving the quality of data, but also hand-written and verbal data converting into digital records. Those expenses simply were not planned in agencies' budgets.

Among adverse consequences [3] emphasizes the fact that the open data programme did not decrease the information divide between developed and developing countries. Causes are in fact that developed countries have good mechanisms for data collecting and experts who use them. Undeveloped countries haven't enough such experts, and data are often of low quality or incomplete. In such countries citizens rarely participate in efforts about using of open data. It is a consequence of lack of knowledge and motivation. Of course, some kind of solidarity toward 'weaker' societies should not be obstacle in development.

Open data empower already powerful corporations and software companies who possess funds and skills for data usage. In other words, rich corporations (such as Lloyds of London, Google, Daily Mail, etc.) are subsidized by open data. It is almost absurd that sometimes government agencies buy its own data, but also there is a danger of corporative monopolization which could shape urban infrastructure guided by interests of profit [5].

Government bodies sometimes don't focus on processes outside their boundaries. That is why the process of opening data they don't see as interaction of government and public. While politicians are focused on stories about economic growth and citizen participation, officials are concerned about issues such as privacy, confidence, quality and data completeness.

Securing personal data privacy is a very sensitive issue. A fact is that most of open data (such as geo or budget data) are unrelated to people. But, personal data from which are deleted names, surnames and other identifiers don't guarantee absolute secrecy. Using only data about sex, date of birth and city in the USA, 53% of the population can be uniquely identified [6].

B. Demand-side issues

The issues are present also on demand side which lags after supply side. Part of causes lies in published data itself. Government agencies often under pressure of data opening have taken passive-aggressive attitude by publishing minimal quantity of almost useless data. At first developers chose interesting data for new applications, but interest fell off quickly. Developers from New Zealand even refused to develop applications with useless data [7]. The second part of causes lies in lack of knowledge how to use those data which are pretty useless, until someone transforms it into something valuable. Lack of technical and statistical knowledge are serious barriers for exploiting potentials. Unknowing tools for effective open data processing is one more obstacle. That is why need for education in fields of statistics has appeared. It is necessary for avoiding erroneous analysis and results misinterpretations. Great Britain's Office for National Statistics, in cooperation with Durham University, trains interested journalists and politicians for using tools which enable data synthesizing from various sources, multivariate data interpretation, and so on.

Process of visualization, as a very important aspect of open data usage, demands special applications which should be easy to use, enable saving intermediate states of processing (due to repeatability), visual display of differences.

Even finding published datasets can be an issue. In New Zealand it is common that every agency publishes datasets on its own website. Since websites differ in design, finding data can be challenging [8].

III. TOWARDS OVERCOMING OF ISSUES

With a goal of relieved use and alleviation of some topical issues, dedicated infrastructure to work with open data are being developed. One of them is ENGAGE e-infrastructure. ENGAGE is based on literature, interviews, questionnaires and workshops, and is developed by Design Science Research methodology (DSRM). ENGAGE supports supply and usage of open data in five categories [9]:

- data provision – except data, it enables metadata acceptance, data cleansing, validation, conversion into other formats and metadata enhancement,
- searching and using data – uses SQL and SPARQL, has module for data demands and version management,
- linking data – manual and automated,
- user ratings – assessment system and quality control,
- user collaboration – user profiles, three user roles: supplier, consumer and coworker.

It is hard to understand how open data are used without previous understanding who are the users of open data and which are their motivations. Research conducted in Great Britain identified six motivation groups [10]: one has the government into focus, the second has focus on technology and innovation, third is focused on making a

profit, fourth deems that government digitalization will have reform of government as a consequence, the fifth represents users to whom open data will be a tool for achieving predefined goals, while the last group consists of entrepreneurs who cooperate with government.

Researchers in [11] identified six types of companies who are primarily oriented to:

- collecting and aggregating data,
- analytics as a service,
- creating data and analysis,
- knowledge discovering,
- data aggregation as a service,
- multi-source mash-up and analysis.

The main goal of a company is gaining a profit, which basically means making larger incomes relative to expenses. Web and mobile applications, as well as various services generate expenses in development phase, but also later on – during operation and changes. So, it is not only a matter of thinking about resources, what to offer, how to create it and who are the users. A company must think about profit side, i.e. how to charge a service. Income model exposed in [12] differentiates five main ways to generate an income:

- subscription model – recommended for frequently used service, but suitable for specialized or (semi) monopolistic information (such as geo maps in large scale) because internet users unwillingly pay online services,
- usage model - means service charge measured in bytes, time or session; suitable for ad hoc users who need access to services (possession is not important to them),
- royalty model – customer will pay a price (in the form of percentage or royalty) only after he creates and sells his product with added value
- free model – suitable especially for government services or products
- hybrid model – such as enticement model in which part of the content is free and the other part is to be charged, or community model (like Wikipedia or OpenStreetMap) where income is generated by subsidiary products/services or by donations.

Besides above-mentioned, income model that can be applied on mobile applications is one-time charge on download. Price admissibility combined with usefulness can generate significant income if application becomes popular. Mobile applications can find one exclusive buyer who will, by further free of charge distribution, create a new marketing channel. Web sites which offer free information generated by open data and have a large number of visitors can attract sponsors or advertisers and ensure income. Start-up companies or individuals can attract partners and find their place on the market. Therefore, the application itself doesn't create benefit, but enables easier finding jobs in the future.

IV. OPEN DATA IN CROATIA

Based on surveys, the Open Data Census has created Open Data Index for more than 100 countries worldwide. Table 1. shows ranking of Croatia and its neighbours according to mentioned index [13]. As we can see, of all Croatia's neighbours, only Bosnia and Herzegovina has poorer standing (no data exist for Monte Negro).

Table 1. OPEN DATA INDEX RANKINGS

Position	Country	Score
26.	Italy	545
28.	Slovenia	540
39.	Hungary	475
49.	Serbia	420
52.	Croatia	405
94.	Bosnia and Herzegovina	205
---	Monte Negro	no data

Information Commissioner (IC), as an independent government body for protection of the right of access to information, highlights following issues of open data in Croatia [14]:

- ignoring requests for information (2/3 of the complaints were about the silence of public authority bodies),
- tardiness of public authority bodies,
- information officers' lack of professionalism,
- overcharging of providing information (when actual material expenses incurred),
- insufficient proactive data publishing,
- low citizens' awareness about the right of access to information,
- uncoordinated regulations (about data privacy, trade secret etc.).

IC has a very important role in resolving complaints related to the right of access to information reuse. It also supervises the implementation of the Law on the right to access information. IC can submit an indictment proposal and can issue a misdemeanour order. At this moment, IC office is underfunded and understaffed. That fact negatively impacts on significant progress in realization of the right of access to information and information reuse.

At the time of finishing this paper (in mid-February) we are waiting for launching data.gov.hr portal (we can expect it in a few weeks, according to announcements). The portal is being developed by use of open source code of British data.gov.uk available at GitHub. It uses CKAN platform for data publishing. At the beginning, about 70 datasets will be published. Like in the USA or Great Britain, the intention is data.gov.hr would be the central portal for publishing open data, so all government bodies and agencies, every city and municipality will be able to put data in one place. This concept has advantages for

both publishers (no need for investment in infrastructure) and users (easier finding datasets). On Croatian open data portal there will be a request form for asking new (unpublished) datasets. The intention is to listen to datasets users' needs to avoid efforts about publishing datasets which will not be used.

Before presenting our findings it is important to emphasize that we analyzed only some publishers – so the list of datasets is not comprehensive. Municipalities and counties were not in our focus at all.

In the current situation, we found that some government bodies have a clear understanding of the need for open data. One of them is Ministry of administration which on its website <https://uprava.gov.hr/> clearly states the terms of use for their 12 datasets published in CSV format. The Croatian public health institute publishes lot's of reports and publications in PDF format, and eight XLS datasets at <http://www.hzjz.hr>. The Croatian health insurance fund published more than 40 datasets in XLS format available at <http://www.hzzo.hr/>. Ministry of economy has recently developed web site for fuel prices tracking (<http://www.min-go.hr>) which helps citizens to find a cheaper gas station. It is expected that raw data will be available in acceptable formats for interested parties soon.

Geoportal is a service which is available for citizens recently. With it, we can see in every moment and examine everything related to cadastral and land registry business. Also via additional links, users can view documents such as proprietary and title deed, and view the history of documents.

Croatian bureau of statistics at <http://www.dzs.hr/> provides national statistics about a large set of various data on a national level. They offer users to download and use their program for viewing and formatting datasets and downloading datasets directly from the website in the format they want.

Another site that offers open data to the public is Croatian waters. Croatian waters on their pages <http://vodostaji.voda.hr/> offer information that is updated every hour. All data are complemented with graphical icons so it is very easily to notice water levels that are higher than normal limits.

Easy access to their data, gives Meteorological and hydrological institute of Croatia (DHMZ). On their site <http://meteo.hr> there are current weather data for a lot of cities in Croatia and Europe. Beside all the weather data that can be retrieved, data on water levels of rivers as well as information about the index danger of forest fires is also available. What DHMZ explicitly offers is dataset in XML format containing data about temperature, humidity, pressure, wind and condition for many Croatian cities.

The Croatian environment agency offers lots of data contained in Environment Information System at <http://iszo.azo.hr/>.

Croatian Automobile Club at <http://www.hak.hr/> provides access to information about road conditions as well as other various services related to road trip with possibility of refreshing rate so it can be said that the data

is updated in real time. There is also a map that offers a view of all information that are essential together with an estimate of travel time, toll rates, the list of restaurants, resorts, hotels and cameras that are installed on highways.

The HAK mobile application offers a very transparent access to all data that are essential for the journey. In addition to what has already been stated, that is available on the website, mobile application offers finding the car if it is parked in a place which is not known to the owner. It also provides reference assistance both on the road and at sea by giving the exact location of which can help in case of a damage or accident with the vehicle.

Croatian Railways at <http://www.hzpp.hr/voznired> provides monitoring service trains in real time. All data can be found on their website and is very easy to find the data that are important for the train journey. The HZnet mobile application provides more readable and easier information about trains in Croatia.

Jadrolinija provides sailing schedule for Croatian ferries across the Adriatic Sea at <http://www.jadrolinija.hr/>. The data is in pdf document so if developers want to use that information, they must convert this data in more accessible format.

Most Croatian bus carriers provide their timetables on their websites. Some of that data is on website as image so they can't be reused. Some of data is in PDF document so it can be reused but it requires reformatting.

Mobile application provides information about domestic and international intercity bus schedule of Croatian bus carriers. We can find departures from each bus terminal in Croatia and phone numbers, locations and other contact information of bus terminals and carriers.

Croatia Airlines provides on its website (<http://www.croatiaairlines.com/>) all information about timetable and flight statuses of planes. There is also a mobile application that mobile users can use for getting information and checking their flights.

All post codes and post offices can be found on the Croatian Post website <http://posta.hr/>. Mobile application of Croatian Post also provides information about post codes and all post offices in Croatia. Users can use their GPS to locate the nearest post office, which is very usable if user doesn't know the area.

All major cities in Croatia provide their city budget open to the public. Most of that data is in PDF or Word document so it is easy to read for the user, but for other purpose, users need to reformat data. All data are in low level of transparency because they show only a rough picture of real costs for some activities and for every kind of deeper data you need to contact city office to give you that data.

City of Zagreb at <http://data.zagreb.hr/> has published 12 datasets. Five of them are related to (un)employment statistics. Two datasets are published in XLS(X) formats, while others are available in preferred non-proprietary CSV format.

Part of research about open government data mobile applications was also challenging. First few applications

we found on webpages of datasets publishers. Then we concentrated on Google Play app store because it has good search engine, some data of our interest (number of downloads, average rating) and last, but not least important, according to [15], Android operating system is present in 69,5% mobile devices in Croatia. Since there is no „open data“ category of applications, we searched terms like Croatia, names of Croatian cities, weather, travel, timetable, budget, health, ecology and many other terms in Croatian language (sometimes also in English). When we found application of interest, we also searched for other applications of same developer. Sometimes we found applications in „Similar“ category. When we exhausted search in play store, we used web search engines and we searched on few Croatian web portals. We stopped when we found 50 apps, because after about 40 apps found, every new was hard and time consuming to find. So we are aware the application list is not final, but it gave us some interesting findings.

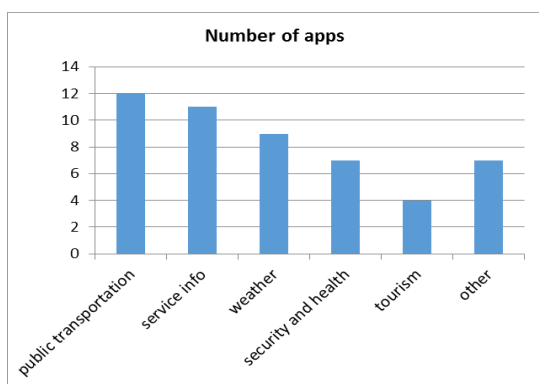


Figure 1. Number of apps per category

Figure 1 shows the number of applications within certain categories. Most of the apps are related to public transportation (bus and train timetables, delays, taxi and tram), various service information (holidays, post numbers and offices, road traffic, ATM machines etc.), weather, and security and health (emergency services, pollen forecasts, bike security, etc.). Our research didn't include classic tourist city guides, so number of tourist applications is quite low. All other apps are related to finance, ecology, city public services and social themes.

Table 2. Downloads and ratings of applications

Number of downloads range	Number of applications	Average grade
<100	4	4,0
100 – 500	7	3,9
500 – 1.000	2	4,6
1.000 – 5.000	9	4,0
5.000 – 10.000	8	4,1
10.000 – 50.000	16	4,1
>50.000	4	4,2

Table 2 shows that one third of apps are downloaded between 10.000 and 50.000, while only four are

downloaded more than 50.000. Two of them are related to weather, one to road traffic and one to, interestingly, finding information about car (type, producer, insurance details etc.) based on registration plates. Only app about earthquakes didn't get any grade, but it is probably because it has been downloaded less than 50 times.

Only one app is not free of charge (it is related to Croatian marines), while all others are free to download. It is hard to say would it be different if Google allowed Croatian developers to sell their apps on Google Play. At this moment, Croatian developers can freely distribute apps, but if they want to sell, they have to open company in one of the countries which are acceptable to Google.

Fourteen apps were developed by individual users, and other by companies and institutions. Almost every app is in Croatian language, many of them are in English (and Croatian), and only a few are in Italian or German language.

V. CONCLUSION

Open data didn't empower economy as it was expected. Publishers of open data meet many issues such as unrealistic goals, expenses, decontextualization and lack of knowledge about what data is interesting to publish. Developers still have to learn how to use this data to create applications that are recognized and wanted in the market. They have to find an income model if they want to succeed. Developing mobile applications is not the only possible activity with open data; there are others – such as analysis creation, knowledge discovery, multi source mash-ups and so on.

Croatian IT specialists are still in a lack of valuable open government datasets, clear licenses and data formats. Despite that, they are quite active. They fight against hard-to-use formats, captchas, misunderstanding, and impossibility to charge app downloads on most popular app store. Portal data.gov.hr will make easier access to open data. About 70 datasets provided on portal maybe doesn't seem as a lot, but the quality (or better to say: usability) is more important than quantity. Developers should ask for opening other interesting data with understanding that publishers cannot always know needs for data in advance and also that publishers sometimes have rational constraints not to open every asked dataset.

REFERENCES

- [1] C. M. Bonina, "New business models and the value of open data: definitions, challenges and opportunities", NEMODE – 3K Small Grants Call, 2013.
- [2] M. Jansen, Y. Charalabidis and A. Zuiderwijk, "Benefits, Adoption Barriers and Myths of Open Data and Open Government", Information Systems Management, vol. 29, issue 4, pp. 258-268, 2012
- [3] A. Peled, "Re-Designing Open Data 2.0.", in Conference for E-Democracy and Open Government, vol. 5 (2), pp. 187-199, 2013.
- [4] G. Eibl and B. Lutz, "Money for nothing – data for free", in Conference For Democracy & Open Government, 2013.

- [5] J. Bates, "This is what modern deregulation looks like: co-optation Data Initiative", *The Journal of Community Informatics*, vol 8 (2), 2012.
- [6] D. J. Hand, "Data, Not Dogma: Big Data, Open Data, and the Opportunities Ahead", *Advances in Intelligent Data Analysis XII*, 12th International Symposium, IDA 2013, London, 2013.
- [7] A. Peled, "When transparency and collaboration collide: The USA Open Data program", *Journal of the American Society for Information Science and Technology*, vol 62, 11, pp. 2085–2094, 2011.
- [8] J. Oh, "Literature Review – Overview of Open Data in New Zealand", University of Auckland, 2013.
- [9] A. Zuiderwijk, M. Janssen and K. Jeffery, "Towards an e-infrastructure to support the provision and use of open data", in *Proceedings of the International Conference for E-Democracy and Open Government*, pp. 275-291, Danube University Krems, 2013.
- [10] T. Davies, "Open data, democracy and public sector reform: A look at open government data use from data.gov.uk", Masters thesis, University of Oxford, 2010.
- and contestation in the shaping of the UK's Open Government
- [11] P. M. Hartmann, M. Zaki, N. Feldmann and A. Neely, "Big Data for Big Business? A Taxonomy of Data-Driven Business Models Used by Start-Up Firms", University of Cambridge, 2014.
- [12] F. Welle Donker, "Public sector geo web services: which business model will pay for a free lunch?", in *SDI Convergence: Research, Emerging Trends, and Critical Assessment*. Nederlandse Commissie voor Geodesie, pp. 35-51, 2009.
- [13] Global Open Data Index: Survey, <http://global.census.okfn.org/>
- [14] A. Musa, "Pravo na pristup informacijama u Republici Hrvatskoj: novi izazovi, nova rješenja", <http://www.pristupinfo.hr/wp-content/uploads/2015/01/01-Anamarija-Musa-Predstavljanje-rada-ureda-PI-240914.pdf>
- [15] Top 8 Mobile Operating Systems in Croatia from Feb 2014 to Jan 2015., http://gs.statcounter.com/#mobile_os-HR-monthly-201402-201501-bar