



# Comparison of Approaches to Publication of Open Government Data in Two Czech Public Sector Bodies

Jan Kučera\*, Dušan Chlapek\*

\* Department of Information Technologies, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, {jan.kucera, chlapek}@vse.cz, (+420) 224 09 5454

**Abstract:** This case study compares approaches to publication of Open Government Data in two Czech public sector bodies: the Czech Telecommunication Office (CTO) and the Czech Trade Inspection Authority (CTIA). The top-down approach applied by the CTO is compared to the bottom-up approach applied by the CTIA. Results achieved by these two public sector bodies are described and the case study is concluded with the lessons learned.

**Keywords:** Open Data, Open Government Data, OGD, Czech Republic, Czech Telecommunication Office, Czech Trade Inspection Authority, top-down approach, bottom-up approach

**Acknowledgement:** The research is supported by the EU ICT PSP project Share-PSI 2.0, Grant Agreement no. 621012.

**Czech Republic** showed its commitment to Open Government Data (OGD) by joining the Open Government Partnership in 2011 (Government of the Czech Republic, 2011). Since then a number of OGD activities have been started at different levels of the public sector as well as in academia and other domains (Chlapek, Kučera, Nečaský & Kubáň, 2014). Although there are some challenges that the Czech Republic needs to face – like the missing official OGD catalogue – some of the public sector bodies have already started to publish Open Government Data (Chlapek, Kučera, Nečaský & Kubáň, 2014).

In this paper we focus on two public sector bodies with the nationwide authority in the Czech Republic that publish Open Government Data, namely, the Czech Telecommunication Office (CTO) and the Czech Trade Inspection Authority (CTIA). Approaches to OGD differ in these two public sector bodies. The goal of this paper is to compare the approaches applied by these two Czech public sector bodies and to present some lessons learned based on this comparison.

## 1. OGD in the Czech Telecommunication Office

The Czech Telecommunication Office exercises state administration in the area of electronic communications and postal services (Czech Telecommunication Office, 2005). CTO's motivation for opening its data up was to increase transparency of the Office and to improve its services (Czech Telecommunication Office, 2013). CTO applied the top-down approach to its OGD when an analysis of the potential datasets for opening up was performed, and a roadmap was established based on this analysis. Initial analysis of the CTO's data involved the following steps (Czech Telecommunication Office, 2014a):

- definition of the openness criteria for the data of the CTO and how these criteria should be satisfied,
- identification of benefits and risks of opening up of the data of the CTO,

- identification of the candidate datasets that should be published as Open Data,
- estimation of the amount of work needed to open up the identified datasets,
- setting up the dataset priorities,
- definition of the terms of use of the Open Data of the CTO,
- definition of the relevant parts of the newly proposed internal guidelines related to the Open Data of the CTO,
- recommendations about cataloguing of the Open Data of the CTO and the metadata attributes included in the catalogue record,
- identification of applications that might help to increase the openness of the CTO,
- specification of the projects required to put the recommendations into practice.

Based on the effort estimates and the risk analysis 50 datasets were selected for opening up (Czech Telecommunication Office, 2013). At the end of March 2014 the first 10 datasets were released as OGD together with their metadata (Czech Telecommunication Office, 2014b). A new section dedicated to the OGD was added to the CTO's website which contains the catalogue of the CTO's open datasets (Czech Telecommunication Office, 2014c). Users are also allowed to suggest new datasets for opening up using a dedicated web form. The CTO expects that the release of OGD will foster development of new applications (Czech Telecommunication Office, 2014b).

## 2. OGD in the Czech Trade Inspection Authority

The Czech Trade Inspection Authority (CTIA) is a public sector body that falls under the jurisdiction of the Ministry of Industry and Trade of the Czech Republic.

*"The CTIA monitors and inspects businesses and individuals, who supply goods to, or sell goods on, the Czech market, provide services or similar activities on the domestic market, provide consumer credit, and operate marketplaces, unless, as a result of special legislation, these activities fall under the authority of another administrative institution." (Czech Trade Inspection Authority, 2013b)*

The Czech Trade Inspection Authority has been publishing OGD about inspections, sanctions and bans since September 2013 (Czech Trade Inspection Authority, 2013a). This data was selected for opening up because it was often targeted in the requests falling under the freedom of information (FOI) legislation and because some of the data had been already published on the CTIA's website (Tajtl, 2014b). Other motivators for opening the data up included improving the prestige of the agency, improving the ways the public is informed about the CTIA's actions, and the prevention of the sanctioned behavior of the businesses and individuals supplying or selling goods and services. According to Tajtl (2014b) the CTIA estimates that during the first year of the OGD initiative it only took up to 30 person-hours in total to open the data up.

Data about inspections, sanctions and bans are provided as six separate datasets in CSV, XLSX ODS (Open Document Spreadsheet) and RDF formats (Czech Trade Inspection Authority, 2013a). The provided data is partially anonymized – data about sole traders is only provided as aggregated data with no personal information included (Czech Trade Inspection Authority, 2013a). Complete records are only available for inspections of legal entities. This approach was taken after a consultation with the Office for personal data protection during which a question of protection of the personal data about the individuals was discussed (Czech Trade Inspection Authority, 2013a).

Because it was difficult to analyze what business activities were the subject of the inspections in the anonymized data, the published open data was later enriched with the NACE classification (Tajtl, 2014a). This allows analysis of the inspections and their results by the standard classification of the economic activities.

CTIA's Open Data has already been reused. According to Tajtl (2014b) the known re-users include universities, one Czech online portal and the data journalists. It is expected that there are other unknown re-users as well. CTIA expects that its OGD initiative will lead to reduced number of FOI requests, improved quality of the data thanks to the user feedback and that it will also help to develop data visualization applications (Tajtl, 2014b).

### 3. Comparison of Approaches

Comparison of the approaches of the CTO and of the CTIA to publication of Open Government Data is provided in Table 1.

Table 1: Comparison of approaches of the CTO and the CTIA

Attribute	CTO	CTIA
Overall approach	Top-down	Bottom-up
Number of datasets	10 released, 50 planned to be published by the end of 2015	6 released by the end of June 2014
Dataset selection	Analysis of datasets held by the CTO, risk assessment, effort estimates, top management priorities	Datasets already published and demanded via the FOI requests
Expected benefits	Increased transparency, improved services, minimizing errors when working with data, easier translations of documents, better understanding and management of CTO's own data, better informed public about administrative supervision, increased value of the data, improved prestige	Improved prestige, lower number of FOI requests, better informed public about administrative supervision, prevention of sanctioned behaviour
Impact	Activity noticed and commented by media and data journalists Improvements based on user feedback	Reuse by universities, journalists and a web portal Application developed (VysledkyKontrol.cz)

As mentioned above, the CTO conducted an analysis of the available datasets in order to identify suitable datasets for opening up. Risks of opening the datasets were assessed and required effort was estimated. Priorities of the datasets were set according to this analysis; preferences of the CTO's top management were taken into account as well. We can call this a topdown approach because the available datasets were identified in the first step and then a subset was selected for opening up.

The CTIA took another approach and started with a limited number of datasets from only one domain. However, the selected datasets result from the main activities of the agency and they are in demand by the re-users. We can call this a bottom-up approach because the agency started with specific datasets rather than with an analysis of the available datasets.

In general the bottom up approach might allow the public sector bodies to start with their OGD initiative quite quickly and gain the necessary experience. According to Open Knowledge Foundation (2012) starting with a limited number of datasets or just one dataset and going quickly might help to build momentum. On the other hand the top-down approach might help the organization to better understand its own data and to better plan and manage the OGD initiative because a roadmap can be developed and the progress can be tracked against this roadmap.

Apart from the focus, the scope of the activities is also different. In the case of the CTO the analysis involved a series of interviews with the representatives of the CTO and two workshops aimed at identification of the potential risks and prioritization of the datasets proposed for opening up. Because the datasets are managed by different departments, collaboration of these departments was required in order to publish the first set of the CTO's datasets and the related metadata. In both of these phases, consultants from the University of Economics, Prague, and the Charles University in Prague participated in the project (Czech Telecommunication Office, 2014b; Czech Telecommunication Office, 2013).

The CTIA's analysis was narrowly focused due to the limited number of datasets, and it was mostly performed by the employees of the agency, although some consultations were also provided by the members of the OpenData.cz initiative (Czech Trade Inspection Authority, 2013a). The Office for personal data protection was also consulted about the potential privacy issues. The smaller scope of the CTIA's OGD activity also required less effort compared to the CTO. The CTIA estimates that the first year of its OGD initiative consumed only up to 30 man-hours (Tajtl, 2014b). Due to the larger scope and the larger number of people involved in the analysis and publication of the first open datasets the CTO's activity took several person-months.

Comparison of these two public sector bodies also shows that the public sector bodies might pursue some similar goals with their OGD initiatives. Both the CTO and the CTIA perform administrative supervision and they see the OGD as a way how to better inform the public about their actions. The CTIA further expects that the publication of their open datasets will help to prevent sanctioned behavior of the inspected subjects. Improved prestige is expected in both cases as well.

On the other hand there are differences in expected benefits too. For example the CTO also expects that OGD will bring some internal benefits and effectiveness improvements like better understanding and management of the data held by the Office.

It is difficult to compare the impacts of these OGD initiatives because the CTIA published its open datasets 6 months earlier than the CTO. However the datasets published by the CTIA are being reused by universities and journalists. Although no analysis regarding reuse of these datasets has been carried out, one of the reasons why the CTIA's datasets attract re-users might be that these datasets had been in demand even before they were published as OGD. In the case of the CTO it is too early to assess the impact because the datasets were published in the end of March 2014 (see (Czech Telecommunication Office, 2014b)). However the OGD activity of the CTO has been noticed by some journalists and some improvements to the published data have been made based on the user feedback.

#### **4. Conclusions**

The Czech Telecommunication Office and the Czech Trade Inspection Authority are two public sector bodies with nationwide authority in the Czech Republic. Both of them have started their OGD initiatives, however different approach was taken by each of them.

The Czech Telecommunication Office applied the top-down approach when it started with an analysis of the datasets available in the Office. Based on the risk assessment and estimation of effort needed to publish the potential datasets as OGD and after discussion of the top management preferences, priorities were set for the candidate datasets. A roadmap for opening up was developed according to the given priorities. At the end of March 2014 the first set of datasets was published and the OGD catalogue of the CTO was launched.

The Czech Trade Inspection Authority applied the bottom-up approach. In September 2013 data about inspections, sanctions and bans was published as OGD. The CTIA decided to start with this data because it had been already requested via the FOI requests and a portion of this data had been already published on the CTIA's website. OGD of the CTIA about the inspections, sanctions and bans are being reused by universities, journalists and a web portal.

The bottom-up approach allows the public sector bodies to start with their OGD initiative quickly and to learn from experience. Selecting the right datasets to start with might be a significant success factor. In case of the CTIA, datasets that had been demanded even before they were published as OGD were selected, and they have already found their re-users. Therefore selecting datasets that are in demand should help to achieve the desired reuse.

On the other hand the top-down approach might help the public sector bodies to better understand the data they hold because in this approach analysis or identification of the available datasets should be performed as the first step. Subset of datasets suitable for opening up should be selected based on the results of the analysis. If the prioritization is performed during the analysis, the OGD roadmap can be developed which might help to manage the OGD initiative.

## References

- Chlapek, D., & Kučera, J., & Nečaský, M., & Kubáň, M. (2014). *Open data and PSI in the Czech Republic*. Retrieved March 31, 2014, from <http://www.epsiplatform.eu/content/open-data-and-psi-czech-republic>
- Czech Telecommunication Office (2014a). *Otevřená data Českého telekomunikačního úřadu: Použitý postup a dosažené výsledky [Open Data of the Czech Telecommunication Office: applied approach and achieved results]*. Retrieved April 10, 2014, from [http://www.issc.cz/archiv/2014/download/prezentace/ctu\\_drtina.pdf](http://www.issc.cz/archiv/2014/download/prezentace/ctu_drtina.pdf)
- Czech Telecommunication Office (2014b). *ČTÚ zpřístupnil první otevřená data [CTO released the first open data]*. Retrieved April 1, 2014 from <http://www.ctu.cz/aktuality/tiskove-zpravy.html?action=detail&ArticleId=11382>
- Czech Telecommunication Office (2014c). *Katalog otevřených dat [Open Data Catalogue]*. Retrieved April 1, 2014, from <http://www.ctu.cz/otevrena-data/katalog-otevrenych-dat-ctu.html>
- Czech Telecommunication Office (2013). *Český telekomunikační úřad otevře svá data [Czech Telecommunication Office will open up its data]*. Retrieved December 8, 2013, from <http://www.ctu.cz/aktuality/tiskove-zpravy.html?action=detail&ArticleId=10859>
- Czech Telecommunication Office (2005). *Basic Information*. Retrieved March 31, 2014, from <http://www.ctu.eu/main.php?pageid=178>
- Czech Trade Inspection Authority (2013a). *Open Data - Databáze kontrol, sankcí a zákazů [Open Data - Database of inspections, sanctions and bans]*. Retrieved December 11, 2013, from <http://www.coi.cz/cz/spotrebitel/open-data-databaze-kontrol-sankci-a-zakazu/>
- Czech Trade Inspection Authority (2013b). *Who we are and what we do*. Retrieved March 31, 2014, from <http://www.coi.cz/en/about-ctia/who-we-are-what-we-do/>
- Government of the Czech Republic (2011). *Usnesení vlády České republiky ze dne 14. září 2011 č. 691 o přistoupení k mezinárodní iniciativě Open Government Partnership [Resolution no. 691 of 14 September 2011 about the accession to the Open Government Partnership international initiative]*. Retrieved November 8, 2014, from [http://kormoran.vlada.cz/usneseni/usneseni\\_webtest.nsf/0/4A3F809FCF5DB30AC125791200296E95/\\$FILE/691%20uv110914.0691.pdf](http://kormoran.vlada.cz/usneseni/usneseni_webtest.nsf/0/4A3F809FCF5DB30AC125791200296E95/$FILE/691%20uv110914.0691.pdf)
- Open Knowledge Foundation (2012). *The Open Data Handbook*. Retrieved August 30, 2012, from <http://opendatahandbook.org/>
- Tajtl, M. (2014a). *Otevírání dat a využití nástroje ODN [Opening up data and use of the ODN tool]*. Retrieved June 20, 2014, from [http://www.cssi.cz/cssi/system/files/all/Seminar\\_CSSI\\_6-6-2014\\_Tajtl](http://www.cssi.cz/cssi/system/files/all/Seminar_CSSI_6-6-2014_Tajtl)
- Tajtl, M (2014b). *Otevřená data o kontrolách ČOI [CTIA inspections data as Open Data]*, ISSS 2014, 7 April 2014. Retrieved April 10, 2014, from [http://www.issc.cz/archiv/2014/download/prezentace/coi\\_tajtl.pdf](http://www.issc.cz/archiv/2014/download/prezentace/coi_tajtl.pdf)

### About the Authors

*Jan Kučera* is a PhD student and a researcher at the Faculty of Informatics and Statistics of the University of Economics, Prague. His research topic is Open Government Data. He is a co-author of the Open data cataloguing strategy and the Open Data publication methodology of the Czech public administration. He co-developed contents of the Czech CKAN instance and collaborated with several Czech public sector bodies on the Open Government Data projects.

*Dušan Chlapek* is an Assistant Professor at the Faculty of Informatics and Statistics of the University of Economics, Prague. He has extensive expertise in development of information systems in both private and public sector. He is a co-author of the Open data cataloguing strategy of the Czech public administration and the Open Data publication methodology of the Czech public administration. He has collaborated with several Czech public sector bodies on the Open Government Data projects. He is a member of the board of the Open Data Forum.