

Comparison of approaches to publication of Open Government Data in two Czech public sector bodies

Jan Kučera, Dušan Chlapek

*Faculty of Informatics and Statistics
University of Economics, Prague*
jan.kucera@vse.cz, chlapek@vse.cz

Abstract: This case study compares the 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. Achieved results of these two public sector bodies are described. The case study is concluded with the lessons learned.

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

1 Introduction

Czech Republic showed its commitment to Open Government Data (OGD) by joining the Open Government Partnership in 2011. Since then number of OGD activities has been started at different levels of the public sector as well as in academia and other domains [1]. 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 OGD [1].

In this paper we focus on two public sector bodies with a 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 provide some lessons learned based on this comparison.

2 OGD in the Czech Telecommunication Office

The Czech Telecommunication Office exercises state administration in the area of electronic communications and postal services [6]. CTO's motivation for opening up its data was to increase transparency of the Office and to improve its services [5]. CTO applied a top-down approach to its OGD when an analysis of potential datasets for opening up was performed and based on this analysis a roadmap was established. Initial analysis of the CTO's data involved the following steps [2]:

- 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 for the Open Data of the CTO,
- definition of the relevant parts of the newly proposed internal guidelines related the Open Data of the CTO,

- recommendations on how the Open Data of the CTO should be catalogued and what attributes should be included in the catalogue record,
- identification of applications that might help to increase the openness of the CTO,
- specification of projects required to put the recommendations into practice.

Based on the effort estimates and the risk analysis 50 datasets were selected for opening up [5]. By the end of March 2014 the first 10 datasets were released as OGD together with their metadata [3]. A new section was added to the CTO's website that is dedicated to the OGD and contains the catalogue of the CTO's open datasets [4]. Users are also allowed to suggest new datasets for opening up using an available web form. The CTO expects that the release of OGD will foster the development of new applications [3].

3 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*" [8].

Czech Trade Inspection Authority published OGD about inspections, sanctions and bans in September 2013 [7]. 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 [10]. Other motivators for opening up the data included improved prestige of the agency, improving the ways the public is informed about the CTIA's actions, and the prevention of the sanctioned behaviour of the businesses and individuals supplying or selling goods and services. According to [10], the CTIA estimates that during the first year of the initiative it only took up to 30 person-hours in total to open up the data.

Data about inspections, sanctions and bans are provided as four separate datasets in CSV, XLSX and ODS (Open Document Spreadsheet) formats. The provided data is partially anonymised – data about individual entrepreneurs is only provided as aggregated data with no personal information included [7]. Complete records are only available in cases where some legal entity is the subject of inspections. This approach was taken after a consultation with the Office for personal data protection [7].

CTIA's Open Data has already been reused. According to CTIA the known re-users include universities, one Czech online portal and the data journalists [10]. It is expected that there are other unknown re-users as well. CTIA expects that its Open Data initiative will help develop data visualization applications, reduce the number of FOI requests, and improve the quality of the data thanks to the user feedback [10].

4 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	4 released
Dataset selection	Analysis of datasets held by the CTO, risk assessment, effort estimates, top management priorities	Datasets already published and demanded by 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	Too early to assess	Reuse by universities, journalists and by a web portal

As it was 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 top-down 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, those covering the main activities of the agency and being demanded 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 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 [9] 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 an 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 publication. Because the datasets are managed by different departments, a collaboration of these departments was required in order to publish the first set of the CTO's datasets and the related metadata. In both phases consultants from the University of Economics, Prague, and the Charles University in Prague participated in the project [3], [5].

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. 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. CTIA estimates it took the CTIA only up to 30 man-hours [10]. Due to the larger scope and 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 publication of their open datasets will help to prevent sanctioned behaviour of the controlled 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 [3]).

5 Conclusions

The Czech Telecommunication Office and the Czech Trade Inspection Authority are two public sector bodies with a nationwide authority in the Czech Republic. Both of them have started their OGD initiatives, however a 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 discussing the preferences of the top management, priorities were set for the potential 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 was already requested in the FOI requests and a portion of this data had been already published on the CTIA's website. OGD of the CTIA about inspections, sanctions and bans are being reused by universities, journalist 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 dataset to start with might be a significant success factor. In the 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 a dataset that is in demand should help to achieve the desired reuse.

The top-down approach might help the public sector bodies to better understand the data that 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 developed that can help to manage the OGD initiative.

6 References

1. Chlapek, D., Kučera, J., Nečaský, M., Kubáň, M. Open data and PSI in the Czech Republic. In: *European Public Sector Information Platform* [online]. March 2014 [cit. 2014-03-31]. Available from: <http://www.epsiplatform.eu/content/open-data-and-psi-czech-republic>
2. Czech Telecommunication Office. Otevřená data Českého telekomunikačního úřadu: Použitý postup a dosažené výsledky [Open Data of the Czech Telecommunication Office: applied approach and achieved results]. In: *ISSS 2014* [online]. 7th April 2014 [cit. 2014-04-10]. Available from: http://www.issc.cz/archiv/2014/download/prezentace/ctu_drtna.pdf
3. Czech Telecommunication Office. ČTÚ zpřístupnil první otevřená data [CTO released the first open data]. In: *ČTÚ* [online]. 31.3.2014 [cit. 2014-04-01]. Available from: <http://www.ctu.cz/aktuality/tiskove-zpravy.html?action=detail&ArticleId=11382>
4. Czech Telecommunication Office. Katalog otevřených dat [Open Data Catalogue]. In: *ČTÚ* [online]. 31.3.2014 [cit. 2014-04-01]. Available from: <http://www.ctu.cz/otevrena-data/katalog-otevrenych-dat-ctu.html>
5. Czech Telecommunication Office. Český telekomunikační úřad otevře svá data [Czech Telecommunication Office will open up its data]. In: *ČTÚ* [online]. 5.11.2013 [cit. 2013-12-08]. Available from: <http://www.ctu.cz/aktuality/tiskove-zpravy.html?action=detail&ArticleId=10859>
6. Czech Telecommunication Office. Basic Information. In: *CTU* [online]. © 2005 [cit. 2014-03-31]. Available from: <http://www.ctu.eu/main.php?pageid=178>
7. Czech Trade Inspection Authority. Open Data - Databáze kontrol, sankcí a zákazů [Open Data - Database of inspections, sanctions and bans]. In: *Česká obchodní inspekce* [online]. September 2013. [cit. 2013-12-11] Available from: <http://www.coi.cz/cz/spotrebitel/open-data-databaze-kontrol-sankci-a-zakazu/>
8. Czech Trade Inspection Authority. Who we are and what we do. In: *Česká obchodní inspekce* [online]. © 2013 [cit. 2014-03-31]. Available from: <http://www.coi.cz/en/about-ctia/who-we-are-what-we-do/>
9. Open Knowledge Foundation. The Open Data Handbook. In: *Open Data Handbook* [online]. 2012 [cit. 2012-08-30]. Available from: <http://opendatahandbook.org/>
10. Tajtl, M. Otevřená data o kontrolách ČOI [CTIA inspections data as Open Data]. In: *ISSS 2014* [online]. 7th April 2014 [cit. 2014-04-10]. Available from: http://www.issc.cz/archiv/2014/download/prezentace/coi_tajtl.pdf