

Russian open data - Current state of mind

Submitted on 14 Sep 2013 by Ivan Begtin

Russian Government as Open Government

I know that in the eyes of many people in Europe, Russia may look like a closed censorship and far from opening up to its own residents. However, strange as it may seem, this is not always the case.

Beginning in 2012, in Russia, a council on public data under the Government commission on openness. Published public data and is actively working on opening more and more information.

I will concentrate on practical examples of Russian openness. Actually we have a lot of them

Examples of data disclosure and it's results

1. Russian government opened up all procurement data including all government tender procedures, contracts, customers and suppliers data. It's about 21 millions of XML documents, 20 GB compressed and 400 GB uncompressed data.

We processed all this data and generated a few reports in 2009 year, 2011 and this 2013. We found that about 1.3% of contracts had errors in suppliers data, about 5% errors with contracts descriptions and so on. Most common problems are errors with organizations identification codes and so on. Our report in 2011 year caused a few meetings with officials from Russian Federal Treasury (state agency responsible for procurement disclosure). Result of these meetings is that treasury launched own project to monitor data quality and they fixed most critical errors.

Our report available here - <http://www.slideshare.net/ivbeg/quality-report-2011> (in Russian). If you need it for translation I could send you MS Word document too.

2. This year, altogether with Russian High School of Economy and Association of procurement institutions, we made deep data analysis and research for such issue as "blind procurement".

"Blind procurement" is unique Russian issue and it's all about connection between corruption and data quality.

Probably you know that some spammers use "word distortion techniques" to bypass spam filters. They replace letters with digits that looks like similarly or do mistakes and so on.

So some Russian _government officials_ used similar techniques to hide tender announcements from search engines. Since Russian procurement data and procedures concentrated in only one website - zakupki.gov.ru, they used limitations of it's search capabilities and used "word distortion techniques".

We found about 10 tricks to hide the data:

- replace cyrillic chars with latin chars that looks like the same
- replace latin chars with cyrillic chars for products with latin name, like "Hyundai" or "Microsoft"

- unreasonable use of dash char "-" inside the word. Like "The ten-der to bu-y mi-lk" instead of "The tender to by milk"
- large amount of mistakes. Up to 10 mistakes per sentence
- and so on

After our research and report published Russian prosecution agencies and procurement control agencies launched audit of all government bodies involved in such activities.

Actually this report is similar to our previous research that we did from 2009 to 2012 years, but this time we found about 12 000 cases of "word distortion techniques" so here is the result.

More info (in Russian) here -

<http://naiz.org/upload/iblock/faa/faac377dabf5757d79a599715b888f81.pdf> and here - <http://xn--80abeamcuufxbhgound0h9cl.xn--p1ai/events/5508098/>

3. Moscow city government published about 170 datasets with geo coordinates at Moscow opendata portal - <http://data.mos.ru>

After quick examination our colleagues from Russian OpenStreetMap community found that data has many errors and mistakes. Actually it was list of about 22 issues with data including errors with CSV files, wrong geo coordinates and so on.

They published research here (in Russian) - <http://gis-lab.info/qa/data-mos.html> and quite soon most of issues were solved by Moscow state officials.

Bad side of Russian openness

We have two non-functioning of government departments - Russian archival agency and the Federal Statistical Office. The first completely forgotten not only about openness, but also on the basis of their work. In particular, they do not put any effort into the preservation of digital archives. The second avoids the disclosure of qualitative statistics. Especially statistics at the municipal level - it is almost impossible to obtain in a suitable form for any use.

We fight for statistical data for a long time and I hope that this year we will get at least basic municipal level

Future is coming

Russian openness is different from European countries. It's much more bureaucratic and right now we have first government law for Open Data and federal government forces regional governments to publish all available data as open data.

At the same time officials are not really very motivated to open the data and we have to fight for any sensitive data like officials statistics and so on.

But for now about 5 Russian regions created open government portals and 3 cities have open data on official websites.

I am absolutely sure that next year open data portals will appear in every Russian region and most of major Russian cities.

Best Regards,

Ivan Begtin

Director of Russian NGO "Informational Culture" (<http://infoculture.ru>) and Ambassador for Russian Federation of Open Knowledge Foundation