Report on the Kazakhstan Open Data for Business Roundtable

NITEC

April 16, 2015

CONTENTS

1. Kazakhstan's open data history and current status

Open data does not refer to just disclosure of information on government agencies' activities, but is a technology tool that enables catalyzing the public and allowing businesses' and common advocates' making use of this original data in multiple projects, as well as performing civil control over authorities' activities.

Governments capture or produce multiple data for their own purposes or based on their own activities. Data may deliver financial benefit where it helps generate policy, distribute resources, or fulfill operational management and optimization. Furthermore, there are non-financial benefits as well, such as meeting legislative and other requirements for transparency and accountability. Generally speaking, if capturing and usage of data did not bring benefit to government, the data would not be captured at all.

Many countries are implementing open data initiatives actively, with one of the key reasons being the great opportunities for economic growth, business innovations and creation of jobs. Experts share same opinion that open data's economic potential is quite considerable, though there is still no uniform methodology to evaluate it. For instance, the UK official research of open data 2013 estimated the direct benefits from open data at £ 1.8 billion, and with indirect ones included as well, (e.g. time saved by virtue of more precise transport information) this amount would be £ 6.8 billion.

Open data use and development would allow obtaining huge economic effect, both on a national scale and company-wide (or project-wide). By the analysis results, Mikhail Abizov, the Russian Federation Minister for Open Government, noted that through disclosure of information in open data format, web-based economy may grow to 3% for 2 years. The social effect will be the creation of new jobs and improvement of quality of life. According to McKinsey research, the open data effect worldwide is estimated by international experts at $ 3 to 5 trillion.

Under the UN methodology for estimating the e-Government performance, open data is one of the key indicators.

There is also the Open Data Index which is drawn up based on estimating the efficiency of governments' activities for enabling the society's access to information around the world. The Index allows civil society's more effectively monitoring the governments' fulfilling their promises as to opening the data. The rating involves 97 countries, and Kazakhstan has got big chance to take one of the leading positions in this rating.

Open data is a fundament the entire Open Government system should be built on, which includes Open Laws & Regulations, Open Dialogue, Open Budgets, and Open Petitions.

Prior to implementing open data in Kazakhstan, the government's preparedness for open data was analyzed based on World Bank methodology. 7 aspects were reviewed, such as: leadership, legal framework, institutional readiness, data availability and demand for it, financing, ICT infrastructure and skills.

The analysis results showed that the most critical of these aspects was legal framework. Therefore, amendments were made to Draft Law "On Informatization", stipulating that government agencies must publish data in machine-readable format according to the list agreed upon by Ministry of Investments & Development. The Draft Law is currently being considered by Mazhilis, with adoption thereof planned for the next year.

The next step was the launch of beta version of open data portal in 2013. By 2014, by virtue of formal requests for data to be provided by government agencies, as well as through independently transforming the data to machine-readable format, the portal data had amounted to more than 800.

In the year-end 2014, a new version of the portal was launched, and also the data structure was optimized, i.e. integrated into datasets. So therefore, the number of datasets is today 217.

On the portal, the Application Programming Interface (API) is also available that enables accessing the data through queries. The portal is accessible at data.egov.kz.

An essential area of open data use is the creation of applications and services based on it, the delivery of services using such data, as well as the usage of up-to-date data for research and for other purposes. For instance, the open data portal was integrated with the systems, such as "Residential Address Register" and "Integrated Regulatory/Reference Information". And this data may be used, for example, in searching by Registration Code of Address, checking a street name, searching for the closest Citizen Service Center or pharmacy, etc. These services are available on the portal.

Also, accessible on the portal are the mobile apps that were presented at the contest among students in October 2014. The winner was the "Kindergartens in Astana" mobile app which allows not only finding the closest kindergarten, but also rating it and leaving feedback.

In February this year, in collaboration with UN and MID of Kazakhstan, the Open Data for Government Agencies Roundtable was held where the idea of open data was explained, as well as its objective, importance, and possibilities of use. The discussion resulted in multiple recommendations obtained, of which the basic one is cooperation with private sector for identifying the priorities in the requested data.

2. Review of open data to be used by private sector

In what way private sector may use this data? It can view data, analyze it, process, obtain some outcomes and use them in its activities. Private sector may also use open data, use in-house data, lay this data on one another, use third-party systems – and as a result, it will get information that can be sold and through which a business may be optimized.

The availability of open data allows:

● to public organizations – exercising control over government agencies` performance,

● to businesses –creating new information products and develop the existing ones,

● to programmers – receiving "the fuel for ideas" in creating new software,

● to citizens – making use of existing applications for increase in personal comfort,

● to government and local agencies – engaging citizens and businesses for creation of new products that contribute to saving public money.

Developers can design useful applications and services using open data and API (Application Programming Interface).

Let us look into some example companies worldwide that are getting huge revenues from open data:

(Real estate) One of examples of high-growth open data market is the 'Zillow' Company that provides real-time data on real estate market. Its share capital amounts to over USD 3 billion. The Company established an online sales platform for houses and real estate with a view to helping the real estate owners, buyers, sellers, tenants, and agents, mortgage specialists, and the landowners renting apartments and managing real estate to search for and share important information on houses and real estate items, on mortgage and home renovation. The Company's business relies on the database of information on more than 110 million houses in US, including those put up for sale, those for rent and those not currently being sold in the market.

(Agriculture) Climate Corp. acquired by 'Monsanto' in October 2013 for USD 930 million uses detailed data on cropping capacity for the last 60 years, meteorological observations at one million measuring points throughout US, as well as the 14 terabyte data on soil quality (free-of-charge governmental data) to provide applications that help farmers increase their profit through making higher quality operational and financial decisions based on information. The key product is the "Insurance Against Weather Conditions" package whereby insurance is paid to farmers automatically, with no need to prove the damage because of bad weather conditions that would affect their revenues. Climate Corp. founded in 2006 г. by two former Google engineers uses 3 million new data elements of 22 arrays per day, applying advanced analysis methods. The data comes from various third parties, e.g. US National Weather Service that publishes its data for free for re-use.

(Building industry) Buildingeye.com allows easily finding and understanding information on construction and planning activities through displaying and matching what

occurs in your city. The Buildingeye's urban platform focuses on GIS and data visualization, using user-friendly interface for citizens to access information on permits for development, on development plans, on investment projects, government actions and other civil activities 24x7.

3. Agenda and attendees

3.1 Agenda

April 16 – Open Data for Business Roundtable		
9:30 - 9:40	Opening • Welcoming speech by a representative from Communication, Informatization & Information Committee under Ministry if Investments & Development, Republic of Kazakhstan	
9:40 – 10:30	Presentations • "What Is Open Data and Why Is It Important?" • "Open Data in Kazakhstan" (National Information Technologies JSC) (attached hereto as Appendix 1) • "Private Sector Use of Open Data" (The Center for Open Data Enterprise) (attached hereto as Appendix 2)	
10:30 – 11:20	Session 1: Private Sector Priorities	
11:20 – 11:40	Break	
11:40 – 12:20	Session 2: Barriers to Use	
12:20 – 13:00	Session 3: Opportunities and Next Steps	
13:00	Wrapping-up	

3.2 Attendees

№	Company name	Sector
1	KazAlko	Alcohol
2	SAS institute	Business, Analytics, Statistics
3	KazMunaiGas	
4	KazMunaiGas	Gas, oil
5	KAZGIPRONEFTETRANS	
6	Research and Production Enterprise Secretariat of Pharmaceutical, Medical Industries and Health Services	
7	Demeu Educational Family Practice Center	
8	Clinic	
9	National Medical Holding	
10	SK-Pharmatsia LLP	Healthcare
11	Sunkar	
12	Meditsell	
13	Medical Association	
14	Dostar Med	
15	BIKO Publishing House LLP	Publishing
16	IBA Group	
17	NAT Kazakhstan JSC	
18	Astana Innovations JSC	
19	ATScom	
20	Real Time Engineering LLP	IT
21	QB Solutions LLP	
22	Crystal Spring LLP	
23	Keremet IT	
24	NAT Kazakhstan JSC	

25	Kazkomnet	
26	"The Internet Associations of Kazakhstan" Association of Legal Entities	
27	Kasip.kz Business Information System LLP	
28	Zerde	
29	Ministry of Investment and Development	
30	Kasip.kz Business Information System LLP	
31	Intel	
32	Astana Innovations JSC	
33	Chelsy-Astana Real Estate Agency	Real estate
34	Information Analysis Center	Education
35	Information Analysis Center	
36	Agrocredit	Agriculture
37	Chairman of Meat Union	
38	Analytical Centre of Economic Police in Agroindustrial Complex LLP	
39	Gosgradkadastr RSE	Construction
40	BI Group Construction	
41	Gosgradkadastr RSE	
42	Kcell	Telecommunication
43	Altel	
44	Kazakhtelecom JSC	
45	Kcell	
46	Beeline	
47	Kazakhtelecom JSC	
48	KAZNEX Invest	Finance

4. Discussions summary and key points

After the completion of first part of the open-table discussions, attendees divided into 8 groups with 7-8 people in each. There were 3 sessions:

Session 1: Private Sector priorities

Session 2: Barriers to Use

Session 3: Opportunities and Next Steps

During the first section, specific features of each and every activity were highlighted. Entrepreneurs told about their experience of data obtainment from government agencies. Thus, a lot of companies request data from customers or use payable resources. For example, subscribe for a payable lawyers` database and legislative acts` database.

Upon the completion of the first section, businessmen stated which data from government agencies they need for optimization and expansion of their activities.

Healthcare:

- o Rating of healthcare organizations:
  - statistics of maternal mortality, infant mortality, tuberculosis, etc.
  - patient satisfaction,
  - average length of hospital stay,
  - extent of risk-associated lethality – by groups of diseases,
  - indicators of quality – by what percent a patient has been healed
- o Statistical data on epidemiological diseases by regions (e.g. chronology or localization of outbreaks of diseases)
- o Statistics on each clinical/spending group:
  - average length of hospital stay,
  - lethality extent,
  - complications extent
- o Statistical data: quantity/volume of imported medications;
- o The needed medicines and which ones have been procured (in terms of names and prices);
- o Information on guaranteed volume of free medicines;
- o Explanation of medications and analogs in plain language;
- o List of out-patient clinics indicating the available equipment and the number of patients registered:
  - Data on availability of ultrasound investigation devices,
  - Data on human resources, length of experience, professional parameters.
- o Which entities participate in the Single National Healthcare System;
- o Data on how many people are employed in healthcare;
- o Testimonials on specialized doctors, to see real ratings.

Construction/ Real Estate:

- Data on land resources allocation for construction;
- Data on real estate developers (whether real estate developer has a license for fund raising or whether he is in a list of mala fide real estate developers);
- Number of students in schools, kindergartens, colleges and universities;
- Information on utility payments;
- Data from Real Estate Registry as to addresses;
- Budget data on building in terms of areas (construction of public facilities, roads, etc.);
- Plans for new construction ЧЖК.

Agriculture:

- Macroeconomic indicators with breakdown into districts and rural districts;
- Every year, government assigns 250 million subsidies to agricultural sector development. This is controlled by akimats. It is necessary to provide monthly report on assignment of such subsidies;
- Information on available unoccupied land with breakdown into districts;
- Free information on weather forecast by coordinates. Currently, this information is received from KazHydromet for payment.

Data on IT market:

- register of IT companies;
- specialists' experience;
- database of specialists.

Data on taxes:

- Data on payment of excise duties (with breakdown into vodka distilleries (importers) – at registration of consolidating authority supervising this distillery);
- Data on legal entities' tax payments (this data would help verify from the beginning some suppliers' and partners' honesty);
- Tax Committee's data, including the sums of all tax data.

Data on procurement:

- Registers of mala fide suppliers;
- Public Procurement;
- Data on bidding processes;
- Data on results of bidding processes;
- Cost of railway transportation indicating type of railway transport.

Legal framework:

- About "Informational/Legal System of Laws and Regulations of the Republic of Kazakhstan";
- Obtainment of information on laws and regulations published in "Adilet" Information System, for commercial purposes;
- Open data from Legal Entities/Individuals National Databases that are not secret ones (a company's details, line of business, number of staffs);
- All government agencies' directories;
- Status of a letter sent to a government agency (i.e. at what stage is the letter, for instance: received, read, under consideration, etc.).

Statistical data:

- Statistical data for marketing analysis of import/export of consumer goods;
- Statistical data on human resources;
- Statistical data of banks and financial organizations;
- Economic and socio-demographic data;
- Register of enterprises:
  - statistics on credit cards,
  - on crisis data, etc.
  - by index,
  - by geographic maps,
  - Household income,
  - Data on areas of activities,
  - Data on CEO,
  - Data on addresses, locations, contacts,
  - Data on filing for VAT,
  - Data on existing debts on taxes and, possibly, on loans (deliberate non-payers),
  - Data on liquidation/bankruptcy,
  - Data on payment of taxes and deductions to budget,
  - Data on availability of licenses for some activity,
- Data on e-Government as a whole, in user-friendly format;
- Classifiers of goods;
- Database on charity (so as to preclude fraud and to get reliable data).

The goal of the second session is to understand the reasons for which companies do not currently use open data or would be hesitant to use it in their businesses.

Social barriers:

- No chance to receive a consultation;
- Awareness (many businesses were not interested in usage of open data before the Roundtable).

Technical barriers:

- Doubts as to data integrity, i.e. data completeness and being up-to-date;
- The data provided is not exact and does not have a single, standard structure;
- There are multiple different sources that contradict each other;
- Executors' incompetence, corruptness;
- Having to wait for delivery for a long time after which information becomes irrelevant;
- Difficulty in searching for necessary information (no data categorization and structure);
- For healthcare: the issue of accessibility of data on medical institutions and in which government orders this data is involved. There are no identifiers for medical products;
- Technical standardization (format, predictability);
- Licensing policy (for use of data for specific datasets);
- Metadata on further use of open data.

Political barriers:

- Unreasoned closure: government agency's confidentiality;
- Government agencies' fear to show real situation in the area;
- Extra work of government agency's staffs;
- There are no clear-cut definition whether the data is open or confidential;
- Data of individual entrepreneurs (barriers between personal data and open data);
- Bureaucratic barriers. Organizational issues. Businesses often face bureaucratic barriers. Government agencies very often are unable to provide clear-cut answers to businesses' queries and forward these queries to other agencies;
- Legal framework is nor broken down by subject matters; there are no Orders and other regulations on government agencies;
- Government agencies do not trust public organizations and private sector.

5. Key recommendations

In the end of Session 3 Opportunities and Next Steps, the following recommendations were given by private sector (благоприятные возможности и следующие шаги):

It was decided to develop a strategy for defining which data can be considered open:

- First and foremost, it is necessary to identify what data is confidential. Then, under the principle:"What is not confidential, must be open" it would be possible to provide all data that is not deemed confidential, regardless of whether it is in demand or not.
- Develop draft law on open data and make the public aware of adoption of the Law;
- Take account of international best practices in providing open data;
- Centralized managerial body that would define posting information on open data (or under the following scheme):
    - Working body:
        - metadata
        - Projects
    - Managerial body:
        - decides whether the data should be opened or not.

Order development for data management and provision:

- Draw up the regulation on open data functioning;
- Data structuring;
- Interaction with government agencies in order to prevent from disembodied;
- Endorse the template containing the rules for a government agency to provide open data, indicating the timelines for frequently updated data;
- For obtainment and usage of commercial data, signing it with digital signature when receiving, in order that organizations that provide this data be sure that the data is not used by malicious persons;
- Create a form for request for data at data.egov.kz; voting

Open Data access improvement and awareness increase:

- Data opening plan: post on the portal the timeline for opening, and inform on which data and what its status is;
- Open data dispatcher. (The key idea was that government should establish an agency to act as dispatcher for data provided by government agencies. Dispatchers would monitor all parameters the data must comply with, such as timeliness, structure, preciseness, etc. As all data is accumulated in this agency, open data could move to the format of big data, which, in turn, would

open opportunities for analytics. Indeed, some processes and trends may be only seen clearly in big data format;

- Provide the examples of implementing the first big projects based on open data;
- Execution of PR campaign for the project promotion and partial funding;
- Establish a group of consultants in specific industries.

The creation of long-run cooperation with private sector:

- Afford opportunities for public organizations to be given government orders to them by government agencies;
- To the extent practicable, organize events related to data and determine for priority use which data is most necessary at the moment;
- Regular roundtables for specific industries with government agencies' representatives participating;
- Do not separate businesses from government agencies.

6. Conclusion and further steps

Roundtable discussions` results show the need to conduct similar events on a regular basis. As it was noted by attendees, businesses` and citizens` demand in open data will be achieved through the project`s popularization and public discussions at different levels.

In order to provide a complex approach to issues solving, questionnaire surveys, roundtables with businesses, and seminars with agencies will be held on a regular basis.

All data you interested in will be requested from government agencies and published on the portal. Also, hackathons will be held to create new ideas for business, as well as products and services based on open data.