ST. PETERSBURG INTERNATIONAL ECONOMIC FORUM

JUNE 20–22, 2013

New Catalysts for Change

OPEN DATA: TRANSPARENCY WITH A PURPOSE

Roundtable

JUNE 21, 2013

09:45–11:00, Pavilion 8, Conference Hall 8.2

St. Petersburg, Russia

2013

Moderator:

Andrew Stott, UK Transparency Board; Senior Consultant, The World Bank

Panellists:

Mikhail Abyzov, Minister of the Russian Federation

Sören Auer, Professor, Coordinator of the pan-European LOD2 (Linked Open Data) project

Jeanne Holm, Chief Knowledge Architect, NASA

Nikolai Nikiforov, Minister of Telecom and Mass Communications of the Russian Federation

Front row participants:

Ekaterina Aksenova, Director, The Strategist Agency

Ivan Begtin, Director, Non-Commercial Partnership 'Information Culture'

Victor Klintsov, Deputy Director of Institute of Information Technology, National Research University 'Higher School of Economics'

Victor Panin, Deputy Chairman, All-Russia Society for Protecting the Rights of Consumers of Educational Services

Sergei Plugotarenko, Director, Russian Association for Electronic Communications (RAEC)

A. Stott:

Good morning, everyone. I am Andrew Stott; I was formerly with the UK Government, setting up their Open Data initiative, and I still advise them. The UK Government is a bit like the Hotel California: you can check out, but you can never leave, so I was taken on as an advisor. I am also now a Senior Consultant to the World Bank, working on open data globally and, indeed, here in Russia.

On the panel we have Minister Mikhail Abyzov, the Minister responsible for taking forward open data in Russia; Sören Auer, who is working on the semantic web and LOD2, Jeanne Holm from data.gov in the US; and I am hoping that Minister Nikolai Nikiforov will join us shortly. He has been a bit delayed, I am afraid, but he is going to speak a bit later.

Each of us is going to talk for a maximum of 10 minutes, and we have got a lot to get through, so I am going to be very strict, even with myself, and then we have got some discussants in the front row who will just say a few words, and then we will take some questions from the audience.

I thought I would just start by saying a few words about what open data is, and why it is important, and why are we talking about it at the St. Petersburg International Economic Forum. For that, the title may be a bit misleading, because open data is not just about transparency: it is about a number of other things, as I hope to demonstrate as we go into the talk.

The basic model for open government data is this: governments collect and generate a vast amount of data, and they use that for their own purposes: for designing public services, for working out priorities for budgets, and so on. Quite a lot of data are generated just through their operations, data about interesting things: things that matter to people. Hospitals, schools, crime, the environment, the performance of government itself, and information about transport. Just because the government has used the data for the reason it was first collected, and the reason that justified the collection of that data, does not mean that other people cannot have bright ideas about how that data can be used; ideas, indeed, that the government had not thought of when collecting that data.

So open data is about making that data available for others to use in innovative ways. That can benefit the economy: people can make money out of that data, can build businesses on that data, can provide innovative services to citizens and to other businesses. It can be used to improve public services, to create consumer pressure on local public services, to complement the improvements that government ministers are trying to make. It can also be used to increase government transparency.

Now, this really started back in about 2007/2008, but the official starting gun was fired by this man, Sir Tim Berners-Lee, the man who invented the World Wide Web just over 20 years ago. He gave a talk called, 'The next web: the web of data', about how important freeing data was to this next generation of the web. He fired this starting gun in February 2009, and now there are about 260 open data initiatives by various governments, regional governments, and municipalities worldwide.

The amount of data being released has also grown steadily. This is the performance on data.gov.uk, the UK Government open data site, which shows that over the three and a half years that it has been running, the amount of data has risen by about nine times from modest beginnings – there is just more and more data coming out.

That data is being used for all sorts of interesting and economic purposes. It is being used to provide business intelligence. What is the government buying, who is it buying from, what price it is paying? And that is stimulating better competition in the government marketplace. It is being used to deliver better information services to the public. Certainly, in my country, the government does not always build the best websites. So other people are innovating new ways, new applications, to get information to the public. The economic value of open data on the London metro system is now 58 times the cost of releasing that data, measured on the same strict criteria that apply to other public transport projects.

It is being used for big data analytics, analysing, for instance, all the prescriptions issued by all the doctors in the UK, to find out what would happen if they applied the latest medical knowledge, and what savings there would be for the National Health Service.

It is about business efficiency, getting data from each municipality on where the roads are being dug up. This is a company called Elgin, which supplies that information to logistics firms, to trucking firms, so that they can avoid areas where there is going to be congestion. Not just where the congestion is right now, but where congestion is going to be tomorrow, or next week, or next month.

It is also about better risk assessment: getting more accurate weather and topographical data to model flood risk, and particularly catastrophic flood risk, much more accurately.

This week, open data has, I think, come of age, because it has been on the agenda at the G8 Summit in Northern Ireland. The eight leaders there signed an Open Data Charter, which established five principles.

The first principle is that government data should be open by default. The question is not why should a government agency open the data, but why should it not open the data. It should be opened unless there is a very good reason not to.

Secondly, quality and quantity. The data needs to be put out there, but governments must also work to improve the quality of their data, and I will come onto that in a few moments.

The data should be useable by all: it should be freely available for any business or any citizen to use for any lawful purpose.

Data should be used to improve governance, and there is some well-established research which has been done by my colleagues in the World Bank showing the link between transparency and better governance and economic performance.

The data should also be available for innovation, for businesses, for entrepreneurs to think of absolutely new things to do with it, things that the government had no idea the data could be used for. For instance, in the UK, I showed that business intelligence example. The UK Government released that data, thinking it was about transparency: it would allow the public to see where the money was being spent. And yet someone has built a new, very profitable business out of that.

On economic growth, the European Commission's Neelie Kroes has described data as the new gold, and the economic benefits of an open data policy for the European

Union have been estimated by the Commission as EUR 40 billion per year of new business activity, and a total economic benefit to the EU, once you take into account time-saving, downstream saving by logistics firms and the like, of up to EUR 140 billion a year.

That business opportunity is not just for IT firms. It is for businesses optimizing their operations, optimizing their risk management, and it is for information-rich businesses providing new and innovative services. It is clear that some of this data is particularly important to the economy. Data such as maps, addresses, weather, company registers, and other official registers really provides the framework for the economy. In the UK now, we are starting to call that the National Information Infrastructure, and it is being seen as a national asset.

Secondly, to take forward open data, we need to build the ecosystem. The value of data is when it is used. Data that just sits on a website, which nobody looks at, is not very useful. So that means building the skills, and the G8 Charter referred to this: economies need to develop the technical expertise and experience to use data effectively; governments need to make sure that data is fully described, that it is not just an untagged spreadsheet, but that there is documentation about what that data is; and government needs to create conditions for innovation to flourish, so data has got to be easy to use, and government has got to provide the framework for that, the incubation services, if you will.

So, the World Bank is working with governments to stimulate this whole innovation and data reuse ecosystem, and working with governments, the private sector, and civil society to make sure the best value is obtained from that data. This includes things like Open Innovation Weeks, which the World Bank has run in a number of countries; things about how to use budget data more effectively; how to develop data journalism, so that journalists can use data responsibly and write data-rich stories rather than scandal-rich stories.

Jeanne Holm will talk about how data is being used in the health sector in the United States, where there has been a fantastic take-up of interest.

And finally, technical excellence. As the G8 have said, quality needs to improve. In the UK, official records record over 20,000 men who are expecting babies, who are pregnant, who have subscribed to midwife services. So there are some flaws in that data. One of the excuses I often hear is "we should not release the data now: we should get it right first". But that is often heard from people who have had the data wrong for several decades. Actually getting the data out there allows other people to help improve it.

In the UK, we released data on 300,000 bus stops – where they were, what services were there – and OpenStreetMap and citizens' organizations submitted 18,000 corrections. So not only is there better data available for citizens, there is better data available for government, too.

We need to start building that into what Tim Berners-Lee originally called the web of data, and Sören Auer will talk a bit about that and how this starts to build the semantic web.

Thank you. That was a sort of gallop through, I am afraid.

Now I would like to hand over to Minister Mikhail Abyzov, to talk about open data in Russia and the plans to take that forward here. Thank you.

M. Abyzov:

Thank you, Andrew.

Ladies and gentlemen, only last week, right before the G8 summit, we saw the conclusion of a major conference in London, which focused primarily on the provision of open data as a requirement for transparency across all aspects of government. The concept of open data underlies the Open Government national project which was adopted in accordance with President Putin's executive order last year. I would like to tell you a bit about how this project is developing and give you an idea of the new goals that we are setting for its next stage.

Under this project, more than 500 datasets will be released in open data format by July 15, 2013. This is quite a lot, if you take into account that up until now the work has been done haphazardly and there has been no regulatory system. The

Government Commission's Open Data Council will oversee the implementation of this plan and coordinate the activities of the key participants. This is not a bureaucratic body, but rather a working group that brings together representatives from the Government, representatives from civil society who are interested in access to timely, high-quality data in up-to-date formats, and representatives from the Internet industry, who are analysing the opportunities offered by open data technology for Internet projects.

The key issue at the beginning of the project was the preparation of the legal framework. For a long time the term 'open data' remained undefined. Two weeks ago, the President signed the relevant amendments to Russian legislation, and based on the work done by the Open Data Council, a law was adopted that regulates the basic technical parameters and terminology, so that we now have a clear idea which information constitutes open data. Roadmaps have been prepared for the 27 key departments to help implement the President's executive order by July 15, 2013 and the goals for the next stage have been set out.

Since the state is collecting vast amounts of information, it is important to establish priorities for the open data project. The issue of priorities was widely discussed at the conference in London. Here is the solution that we found. A platform is currently being developed, sponsored by the Council, which will allow the balance of supply and demand for open data owned by the state to be determined in real time.

While the general concept of open data is being developed, several private projects are also underway. First of all, there is an open e-budget project, something which civil society and business have long had an urgent requirement for. With the help of the Open Data Council, the e-budget concept has been adopted and the Ministry of Finance of the Russian Federation has created a new information architecture. We expect the project to be launched by January 1, 2014. Then there is a project on crime statistics, which is already underway and is being handled by the Office of the Prosecutor General of Russia. Criminal statistics, including geolocation information, are one of the most interesting types of information and some of the most often

requested by society. The next large-scale project is the creation of an electronic facility for government procurement.

Last year, Moscow began to implement the open data project and the regions are catching up with it. Moscow and Kazan are in the lead for now. Moscow has released over 160 datasets. Special applications are being developed to help people to work with open data, which has a direct impact on everyday life. More than 1.3 million people have used the Moscow information resources, primarily the housing and utilities, education, and health applications.

What are the next objectives? First of all, to launch a single open data portal. The Ministry of Communications and Mass Media has prepared the concept, we reviewed it at the Council meeting this week and approved it in its entirety. In addition, we need to train federal employees and to carry out an open data public awareness campaign. Open data, as Andrew mentioned, is a big help to journalists and the service sector. We plan to release many more datasets next year.

Last week, the annual results of the rating of G8 nations in terms of their work with open data were released. Unfortunately, the Russian Federation was rated last. The reason is simple: we have only been working on this project for a short time. Our goal is to make a breakthrough and we have reason to believe that Russia can become one of the G8 leaders in this area. Thank you.

A. Stott:

Thank you, Minister.

Now I would like to invite Jeanne Holm to talk about her experience with open data in the US and what she sees as the next priorities for open data. Jeanne.

J. Holm:

Thank you so much.

One of the things that I wanted to focus on is the idea of why we are trying to release data: it is really to drive economic growth and citizen services, and particularly to see how we sustain that over time.

When we do this, part of the challenge that we are trying to overcome is when we convene as government organizations, when we convene developers and private ventures and others together, how do we make sure that that is a sustainable option?

We need to think of open data as an ecosystem. So just as with our oceans, when we look out here across St. Petersburg, we see all of the things that are necessary to create a healthy environment in oceans and rivers, that same kind of inter-relationship is necessary around open data.

If we look at that, what are the things that are necessary? Here in Russia, we have seen huge progress in opening up data at city and regional levels, and now we are all anxiously awaiting, next month, the new Open Data Portal, so we are very excited about that.

Certainly data is a fuel for innovation, but on top of that, we have to look at the technology. Sören will be talking soon about the idea of linked open data, which is really a forward-looking idea about how we can use some of these technologies to be very innovative and to be able to build on the power of the Internet. But let us bring it down to just thinking about the technologies that we use today. You can start open data with simply a website and links, making that data discoverable and accessible, but you can also build on that technology in some interesting ways that I will talk about.

Then we need to think about the participants. It is not just government agencies, although that tends to be a focus as you get started: who is it that is going to be able to share that data? But we also want to look at who is going to consume that data, who is going to get secondary use from that data, how do we measure whether or not that is actually making a difference here in St. Petersburg, or here in the Russian Federation?

As we look at these ecosystem components, we talk about the types of data that governments are releasing that are actually making a difference. Governments release all kinds of data. In the United States, we have over 400,000 datasets on data.gov, but of those datasets, to be honest, some are not used very often. Some

are used a lot. And this was not necessarily something we were able to predict at the beginning.

So, when we look across cities, regions, governments at the federal level, the kinds of data that are making the biggest economic impact are in health care, public safety, research, scientific data, and transparency. I will talk a little bit more about health care, but when we think about it, these are the things that our citizens care about. These are services that they would buy, if they could help make their children safer, if they could find new ways of getting better access to health care.

As a business owner, I would want to know what kind of research could drive new innovations in my manufacturing company, and what kind of scientific data could drive new innovations in my research and my services. Certainly, and the Minister was talking about working with data journalists, when we think about transparency, advocacy groups and data journalists are really looking at what kinds of data the government is releasing related to transparency.

So, how can our citizens and our journalists help us as a government to be more transparent but also to be more efficient, and to give them the kinds of services that they want? Let us look at these a little bit; let us think about those technology choices.

The first thing, of course, in technology, is to create a site that allows the discovery and download of that data, which can be, again, very simple. But as the datasets grow, you will find that the need to make the process sustainable means that your ministries and agencies have to be able to manage, modify, and maintain their data. So, over time, they will have new datasets, either related to the data that they have released before, because maybe it is an annual survey that they release, or they could realize based on feedback, as we have in the United States, that sometimes there are errors in that data, either errors in the way the data file is constructed, or people want it in different formats, or maybe the data is not clearly understandable to the public, because we use some strange government language. So we need to make sure that our ministries and agencies are able to manage and maintain that

data, and update it and modify it over time, which means that a data catalogue capability and a data management system are necessary.

I have been to hundreds of events over the last couple of years with developers, to try to help them use government data. What they really want is not just the ability to download the data, but the ability to dynamically get to that data via APIs – application programming interfaces – and by web services. So not only do they want us to release that dataset, but they also want the data to be released in ways that are going to make their application or their business analysis very, very real-time and dynamic.

One of the products that several of the governments have come together to provide is something called the Open Government Platform. This is available as free, downloadable open source. It brings in best practices from India, the US, Canada, Ghana, and the emerging work going on in Rwanda as well, and many other nations. This is one kind of activity and one kind of platform that would be available to help people.

The participants in this ecosystem include everybody from the government through to developers and businesses, citizens, advocacy groups, data journalists, and researchers as well. One of the interesting points in time that we are at now is the point where big data and open data are starting to converge. When we think about big data, it really means that there is huge potential for business innovation and business impact.

As we look at the ecosystem, part of what we are also thinking about are the policies that the government has had. The Minister was talking about some of the amazing work going on with the Open Data Council, which is really a starting point for an ecosystem, and also the idea of supply and demand, which can be very difficult to manage sometimes, because citizens will want lots of different data, and they will ask for things that you do not even collect as a government. So it is not that you would not release it but simply that you do not have it, or maybe it is private information about citizens that you cannot release because you do not want to release that generally to the public, or national security data. So, in some cases, the

policies have to respect the need for the government to keep some things private, while still understanding that they are trying to default to this idea of all data having the potential to be open.

So we find that these policies, certainly in the US, and across the globe, sometimes come from central government, but they also come out of ministries, local governments, and advocacy groups. These policies get pulled together into this ecosystem, to be able to show the best way of sharing the data and the best way of being able to manage it.

In each case, when we think about the release of data, we go back to businesses. How can businesses use this data to achieve success? They use data, in the end, to create economic growth. Now, sometimes they are doing it for a social good, but the bottom line is that they are not going to invest in something that is not going to actually be able to benefit their bottom line and make it more profitable. So the outcomes that they might get from open data might help them to create better policies, to find out where they want to invest new resources, how they want to allocate existing resources, applications, visualizations, and other services that they can provide to their customers.

We have documented some of these cases in the US. For example, we release weather data and global positioning satellite data, for free use to anybody in the world, and we have grown a USD 100 billion a year industry based on the delivery of just these two services. They fuel everything from weather forecasting to travel services: almost any kind of service that you buy on the web that has an event location will also talk about weather at that event. And GPS, all of us have one, two, three, four mobile devices on us right now, and all of those are using applications that are GPS-driven. Without those services being freely provided, it would be very difficult to be able to see the kind of innovation that we see.

There are other examples, for instance in the health sector here. iTriage was not developed by anybody in the developer community: it was developed by emergency room doctors who wanted to save lives by getting people to the emergency room faster and by redirecting people from the emergency room who did not need to go

there. So, you enter your symptoms, it uses government data about the ratings of hospitals and the way that they treat their customers, along with success metrics for recovery, and it gets you immediately to the best doctor that you need for your situation.

So, as we see this convergence of big data and open data, we are really seeing this huge growth in the potential for businesses to use their own private data, combined with government data, to be able to create economic growth. As this emergence has happened, we want to make sure that we are looking at the potential for how we can apply this here in the Russian Federation, and so we see this huge uptick in the growth of jobs in this sector.

So, by the government releasing open data that can be mashed up or combined with data from businesses that they are not necessarily going to open up to the public, the businesses themselves are going to be able to create huge economic growth.

It is great that you are launching the Open Data Portal here. I am so excited, and I am really excited, too, to see the events that are going to happen around developers building on this open data from the government to create applications for citizen services, and more importantly how businesses are going to create new economic growth here in St. Petersburg and across the Russian Federation.

A. Stott:

Thank you very much, Jeanne.

Can I now turn to Sören Auer to talk about the semantic web of open data and how we really get that data not only sitting by itself, but joining together, so that you get a network of effective data?

S. Auer:

Thank you very much.

It is very nice to be here in St. Petersburg. I used to come to St. Petersburg often. It is great to be taking part in a discussion on such a subject as open data.

As we have heard already, open data has huge potential and there are many benefits to be gained from it. I think open data is not enough: we actually need to do more than just open data. We need to connect this data, to interlink it, because I think the real value of open data can only be exploited if we can connect data from different sources. By connecting data from different sources we can create new applications, unforeseen applications, and new services, new products, based on this data.

This is how this web of data has emerged over the last year, so you see here a growth in the number of datasets which are published on the web and interlinked. This already shows to an extent that there is this global movement of linking data, bringing data from different sources together. We stop drawing this picture here, because there are so many datasets on the web published in structured formats and linked to each other that you cannot even draw it on a map.

It reminds me a bit of the early days of the web; everybody was drawing pictures of how web servers were linked to each other, and I think we are in a similar situation in this web of data, where we are publishing not only data in different formats but data in standardized formats – there are W3C standards – and linking this data together. In the same way that we publish information on the web and have links between different websites, I think we also need to establish those links between data items, and that will help us, in the end, to create value out of the data.

We have these three aspects of what linked open data actually is. We have open data, we have government data, and then we have this intersection, and we have structured data which is standardized and published according to established standards, and when we link this together, then we have this linked open data or linked open government data.

I think it is very important to build on those standards which have been developed by W3C, by the semantic technologies community, by the linked open data community, to facilitate the exploitation of open data in the economic realm, but also in the societal area.

Then, data can really become the new oil, or the new gold, as Neelie Kroes phrased it. I think Russia has a lot of old oil, but it is always good to have different wells of oil, right? I think data and linked data, and interlinking data from different sources can really become a new oil well for all kinds of economies, and we have to tap these oil wells.

So, what does an architecture for open data look like? If you look at old, traditional architectures, we have a data layer, we have an analysis layer or a data management layer, and then we have a presentation layer on top, and this was controlled by one organization, one agency, one publisher. Now, in the new world of open data, we should integrate layers of openness into this architecture so that we have an access layer, so that we can actually access the raw data as well, or standardized data and standardized structured formats. Similarly, we have an access layer above the analysis layer, and on top of the interaction layer. That makes it much more flexible, and allows third parties to create value out of this data, develop new applications and services on top of it, and exploit this data, which was originally published by individual organizations.

I want to show you some examples. One is Open Spending in the UK, which was developed by the Open Knowledge Foundation. It shows that citizens can actually build apps and experiment with applications visualizing governmental spending. This can start a discourse in society, bringing citizens and governments closer together to start a debate about certain things. I think this is a good thing, because then you do not have unexpected surprises later on, if you find out that taxpayers' money has maybe been spent in the wrong way. You can maybe reach some societal agreement on what the priorities are in your society, and you can also identify over-spending or potential for savings, or misuse of money. I think this is a great example which was developed in the UK, but is now in lots of different countries: in Germany we now also have an open budget with different levels of open data.

Another example is education. I think education systems are one of the most important aspects of society, because we educate basically the future of society,

and it does not always work very well. So improving education is always a good thing, and I think open data can contribute a lot. This screenshot is actually an application in Germany where students can rate their professors. This would be facilitated a great deal by open data, if universities, for example, published information about professors, about lectures, about courses, and so on, and then students can basically rank and rate the professors, they can say what lectures they like and do not like, and also give feedback. This feedback can be used to make improvements. The competition between different educational institutions creates more transparency; you can find out, as a student, where it makes sense to study, which courses you should attend, who the inspiring professors are. You are able to identify excellent or also underperforming institutions in education, and I think that creates much more transparency and has great potential.

That is just one example in education, but you can apply similar strategies in health care, so we have similar trends where you can actually rate your doctors, physicians, and hospitals, and that helps people to find out where they can get the best treatment.

Public administration, environment: you name it, there are thousands of examples where this model can be applied, and I think that public administrations, governments, can and should facilitate that, to create this transparency and get citizens involved, and maybe also crowd-source the involvement of citizens a little bit.

There is another example which is quite famous in the UK, Fix My Street, where citizens can actually supply information about where there are problems, what is bothering them, and can give this feedback to public administrations. That also actually makes the life of the public administrations easier, because they do not have to run around and check everything themselves; they get this feedback from citizens. It is a way in which crowd-sourcing can actually help to make them more efficient and effective in delivering public services.

A last example, which has already been mentioned, is public transportation, public services, and Minister Abyzov also said that creating a higher quality of life in cities

is one important aspect of open data. I think open data can contribute a lot there. We already have this global transportation data available, for example, which makes it easier to create applications, but this should not be accessible only to a few monopolists like Google or Yandex; start-ups and small companies should also be allowed to use this. This screenshot that you see here is an Android app, called Öffi in Germany. It is not released by Google and it is one of the most popular apps in Germany, allowing you to check how to get from A to B. It is very user-friendly, even more user-friendly than Google Maps, and it is widely used in Germany. Unfortunately there is not much open transportation data published yet, so the people who developed this app with start-up employees had to scrape this from websites. This, of course, makes it much more difficult for them and I think it is very important to give equal chances to start-ups and big companies alike.

This is also something that open data can do: ensure that not only the big companies, the big players, have access to data, but also the small ones, the start-ups, the citizens, so that they can build new applications on top of that.

So, what has to be done? I think we also have to publish open data in machine-readable formats, in standardized formats, because this also facilitates use by individuals, by smaller companies, of new ideas. If you release open data in difficult to process formats, it is very difficult to build those applications.

One example is the European Commission, which published all of its financial transparency data in a huge xml file – 200 megabytes of xml. It takes several weeks to process this xml file and find out what subsidies went to which companies and organizations. What we did was transform that into rdf, interlinked it with cities, for example, and with other datasets. This means that you can now browse this data and you can actually find out who, in a city, got subsidies from the European Commission within just a few minutes, without needing to be an expert in xml technologies or spending weeks processing this data.

That is, I think, what governments and public administrations should do: they should publish their data in easily digestible formats, in structured formats, which are already interlinked with other datasets.

Tim Berners-Lee has developed this five star scheme, so that is what we see here, and public administrations should strive to add as many stars as possible to their data.

A. Stott:

We are out of time.

S. Auer:

I had some more screenshots, but perhaps I will make a concluding remark.

I think open up data and then the applications will follow – I think that is a great summary. You do not have to think in advance of all the possible applications, but the community and ecosystem of companies, users, and citizens can create applications out of this open data. Thank you very much.

A. Stott:

Sören, thank you.

Finally, I would like to turn to Minister Nikolai Nikiforov, the Minister of Communications and Mass Media for the Russian Federation.

N. Nikiforov:

Good morning, colleagues. In my opinion, we have heard some very interesting stories about how the concept of open data and its practical application in different countries can change the principles according to which state and society interact, create new markets, and, most importantly, improve standards of living. My colleague Mikhail Abyzov has already said that a project based on the concept of open data has been developed in the Russian Federation: it is currently under review and we expect that it will be approved by the Government in July.

We all understand that government bodies and government institutions in Russia will have to undertake a lot work. At the start of such a long journey, it is important to establish priorities and organize the process in an appropriate way. That is

precisely why we have suggested the central open data portal concept, which will help to establish priorities and facilitate the exchange of best practices. On the one hand, the portal will create an opportunity for the Government and society to communicate by collecting and discussing suggestions from citizens. On the other hand, it is a site where government bodies can share their experiences with one another to resolve technological and, very importantly, legal issues: changes will need to be made to many of our laws and government resolutions. We all know that this process will require multiple reviews and approvals, and it is very important that we get through them quickly.

The third aspect is the creation of a sort of ecosystem. I still remember that slide about the ecosystem of the ocean: in my opinion, it is the perfect metaphor for the open data system. It is very important to introduce management tools that will guarantee that not a single government body will be able to avoid implementing the open data concept. We suggest actively using the tool for coordinating expenses related to IT implementation. Our Ministry is responsible for ensuring that the budget funds assigned to the implementation of these projects is appropriate. The total budget that is coordinated via this type of mechanism is around RUB 80 billion, which is around USD 2.5–3 billion. This is quite a large amount and it requires that certain obligations to disclose data be introduced. So our colleagues in other ministries and agencies will have to include technical projects to develop new information systems in their programmes for years to come. We believe that the existence of the concept and the monitoring of its implementation through the budgeting mechanism will enable us to proceed fairly quickly.

In concluding my presentation, I would like to emphasize the key role of strong teamwork between the ministries. I would like to again thank my colleague Mikhail Abyzov for creating the platform on which this teamwork can take place on a regular basis. The experience of our foreign colleagues demonstrates that this is key to the success of an open data concept. We respectfully and attentively assess international practices and aim to learn from the best of them. Thank you very much!

A. Stott:

Thank you very much, Minister.

Now, we have got a few distinguished guests in the front row, and I am just going to ask them to say a few words and pose them a question, and then I will open it up to some more general questions from the audience.

So, first, Juan Navas-Sabater from the World Bank. How is the World Bank helping the open data movement, and what more can the World Bank offer in this area?

J. Navas-Sabater:

Thank you very much, Andrew. I will try to answer this question quickly and in Russian. First of all, I would like to thank the Forum organizers for giving us an opportunity to participate in this session, since open data policy is a key issue for the World Bank. As you know, the World Bank is also releasing its data and this issue plays an important role in our partnership strategy with the Russian Federation. We also work at the regional level. In particular, we are assessing the readiness of Ulyanovsk Region to implement a comprehensive open data system. We are also collaborating with a number of federal agencies. In December, we participated in an international conference on open data and sponsored a number of different seminars. We are planning a new global partnership on open data and we would really appreciate it if the Russian Federation would join this partnership.

In conclusion, I would like to say that we are ready to continue and to expand our partnership with Russia both at the regional and at the federal level and, of course, to offer our help at the global level, to enable Russia to achieve the goal that was mentioned today: to become one of the leaders among the G8 countries. Thank you very much.

M. Abyzov:

I would just like to say a couple of words. The open data concept ensures the interaction and efficiency of many initiatives, including at an international level. At

the conference which took place during the G8 summit, the World Bank introduced the concept of disclosing information regarding the ultimate beneficiaries of offshore companies and legal entities in general around the world. It is open data decisions that form the basis of this concept and of the product that the World Bank is planning to launch in the near future. They cover information on both the companies' ultimate owners and their interactions and cross-ownership issues.

It is obvious that today, it is impossible to implement large-scale initiatives in the global information space without employing an open data concept, which not only improves standards of living, but also helps to facilitate international economic projects focused on healthcare, transparency, and combatting the use of offshore tax havens.

A. Stott:

In terms of this use of data between governments and the open data on this international corporate world that has caught up on governments, I am an advisor to a body called OpenCorporates, which is trying to take these company registers and compare them. We take some globally well-known brands and we find that they have hundreds of subsidiaries in just the 60 company registers we have managed to capture so far.

You have to think, why have they got all of these subsidiaries? What is the point of this? What is the advantage? We can guess, with some of the things that have come out about where tax is actually being paid, and whether it is being paid where the economic activity is. So there is a whole world there that is just starting to emerge as open data becomes available.

We have got four other guests from Russia, from the users of data side, so I would like really for them to talk just for no longer than a minute please, on what they think are the two or three most important things that need to be done in Russia to get the benefits of open data to flow in Russia?

First, I will talk to Ivan Begtin, who has been agitating for open data here in Russia for a number of years. Ivan.

I. Begtin:

Thank you, Andrew. In my opinion, there are three key issues now: commitment, training, and workplace culture. Regarding commitment: government bodies are currently releasing vast amounts of data, but in just a few months they will start saying that no one is using this information and that it was all useless. This means that the project will not function without an active community of developers who will use this data. Many Russian organizations hold competitions for developers, but in addition to these competitions, developers need to have close contact with government bodies. The Government must be ready to publish data in the most convenient formats, and to carry on a dialogue so that their systems can be adapted to satisfy citizens' needs.

Training: we have encountered an incredibly low level of awareness of the concept of open data among federal employees. This is in part understandable, they are busy with their own industry's challenges; but on the other hand, the issue has been discussed at the federal level for over a year now and for many years at the international level–. So we urgently need to introduce training for government officials and business representatives; we need to introduce the subject into the education process in all forms – from open data schools to career development courses.

Third: the lack of ethics. We see active opposition on behalf of government officials. If they purposefully treat their task as simply a formality and interfere with machine-readable data, it is my opinion that they should be fired or severely punished in some other way.

A. Stott:

Thank you.

M. Abyzov:

I would like to comment. This is a good triad, I would rephrase it as: training, enforcement, and pleasure. I would like to comment on what Nikolai Nikiforov and Ivan Begtin have said.

Regarding training: the Open Data Council proposed a project (and the Open Government Commission approved it) according to which open data classes will become a part of federal employee training starting this year. Thus we will provide federal employees with the required skills and will improve the operations of state institutions in that regard.

Then there is enforcement. Nikolai has already said that we will use all available tools to encourage government bodies to release information in open data format. One such tool is the prioritization of spending on automation and on issues relating to the release, storage, and processing of data. Should the state encourage the development of applications and special products through federal investment? Around the world, governments have financed development at the first stage of the process. For the time being, we have not made a similar decision, because we would like to motivate the industry to propose its own solutions first.

Finally, pleasure. What the state can afford to do is to create special funds for start-up development. Such funds exist in many countries, they provide both seed financing and financing at the advanced stages of project development. We have several development institutes, including Skolkovo; the creation of a start-up Internet fund under the Agency for Strategic Initiatives is now being discussed; we have the Russian Venture Company, and we hope that they will all play an active role in implementing the open data concept and creating Internet products that are useful to society and to economies based on this system.

A. Stott:

Thank you. I think my old Minister in the United Kingdom, Francis Maude, said that he was always being told that there needed to be culture change, and his model was: if you enforce hard enough, and change behaviours, then eventually the culture changes as well, whereas if you wait for the culture to change, you will be

waiting forever. So, I think that point about enforcement, about measurement, about comparison, is very important.

Can I now turn to Ekaterina Aksenova to talk about what you see as the priorities for the next stage?

E. Aksenova:

I would like to bring your attention to something very important that I have noticed as a result of many conversations with businesses about open data. The meme that data is the new oil and the new gold gives the wrong idea, to a certain extent. Everyone has decided: all right, let the state create a new oil well and a new Alaska for us, and we can take it all and earn some money. But the truth is that however well all the ministries and departments might work, they cannot create a high-quality open data infrastructure – a working system and not a dead portal – on their own. Many countries have created open data portals, but 90% of them do not work. An effective infrastructure can only be the result of a private–public partnership (PPP) and the portal that is designed now should from the start be a place for PPP, where businesses would not only take, but also give. What I have in mind, in particular, is the option for socially responsible businesses, whose operations are important to all citizens, to be able to upload data; an option to share access to large and expensive data processing platforms, because when you are working with seriously big data, and not an Excel spreadsheet, you need serious tools and they cost a pretty penny. There is a whole range of other types of private–public partnerships without which no ocean is possible.

A. Stott:

I think that is very important, to see the portal as really a platform for communication between the suppliers of data and the users of data, and that means community facilities, the ability to share information about the data there, and not just somewhere where people simply go, collect data and then go off. Some of the best

portals actually have a showcase of the applications that have been made using that data, which stimulate thought.

M. Abyzov:

There have been many discussions about the concept of a single open data portal, and in the last two months we have probably analysed the entire international experience. I am grateful to my colleagues Minister Nikiforov and Minister Belousov: both of these key ministries were at the forefront of the effort. We made a decision which I believe to be very balanced. The Commission Chairman approved it in its entirety. The gist of it is that data owners, namely our ministry, administrations, and relevant federal services, are responsible for open data and its quality. Open data has two key parameters: quality and access speed – and the closer the site is to the data owner, the better these requirements are fulfilled. On the other hand, the distribution of datasets across a huge number of federal agencies does not let you see the entire picture. So we have decided that the single portal will act as an integrator. Without taking the functionality away from the owner's site, we will integrate the data and register it in a single system.

This resource must be implemented as a public–private partnership. I am hoping not just for ideological participation from the Internet industry, but also for investment in the development of this resource. The portal will become a site first of all for training, secondly for start-ups, and thirdly for developing a new strategy. This is the thinking which forms the basis for the concept, which was presented to our colleagues from the Ministry of Communications and Mass Media and the Ministry of Economic Development by the Open Data Council. I would like our colleagues from the Internet industry and the community organizations, whose representatives are here and who make a lot of use of open data, to actively participate in this effort.

A. Stott:

Thank you. Moving quickly to Sergei Plugotarenko, if you could give your views on what are the most important things please.

S. Plugotarenko:

Thank you very much for letting me speak. I am going to talk about connecting the Internet industry to the open data concept. I represent the RAEC – the Russian Association for Electronic Communications – which works with three components: the economy, users, and technology. Russia has the largest number of Internet users in Europe. This means that the market is enormous and that it is growing. We have established five principles on the open government site that will enable the economy and business on the Internet to grow at the same rate, and maybe even faster, than they are now. The open data principle is one of those principles. An in-depth analysis has shown that opening up data and building business based on such data is as important as the development of infrastructure and education, the investment environment, and legislation. If you consider the Internet not just as a business, but also as technology, then it becomes obvious that it is a conducting medium for all ideas and processes that can work with open data, help to store and process it, provide access to it, and offer services based on it. Back in February of this year, the RAEC conducted its own survey among businesses and Internet companies and here are its key conclusions: business considers the open data concept to be popular and sufficiently promising for new services and businesses. The entrepreneurs compiled a rating of the most sought-after types of data based on which business can make money. Expectations are also high for the growth of applications and services which can be built on the basis of open data. At the same time, Internet companies are convinced of the need to promote the idea, to involve the developers, and to be in constant contact with the industry; and the open government site seems to be ideal for that.

But the key recommendations concern the business climate and legislation. If we continue to adopt incomprehensible laws to regulate the industry, we will only slow down the development of the Russian Internet. In the meantime, there are things we can be proud of. I have already mentioned that we are the first in Europe in terms of number of users; we are sixth in the world, and we have our own services – we

cannot lose that. And open data, all other things considered, will certainly become a good business platform. Thank you.

A. Stott:

Thank you. We are just about out of time now, but I know Victor Klintsov of the Higher School of Economics has been sitting there very patiently, so Victor, a few last thoughts from you, please.

V. Klintsov:

I will voice the view of the World Wide Web Consortium, which is promoting the concept of smart data that can be better and more easily used by business. In the nine months since the launch of the Russian open data project, we have reached the stage where datasets have been distributed, as Minister Abyzov rightly said. But in a few months, it will be time for someone to gather these datasets. If the state assigns dataset collection only to business, if it does not attach any strings to them, then not all the datasets will be visible to business: so this dataset went around the corner, how do you pull it back and use it? The state should be concerned about the quality of data, and the highest quality data is linked open data that is easy for businesses to work with. Otherwise, we will have these datasets that have been scattered and forgotten by the state and business will never see them. That is my opinion. Thank you.

A. Stott:

Thank you. We are right out of time, but I do not know, Minister, whether you wanted to say any final words? Jeanne, any final thoughts?

J. Holm:

I am just really delighted about the progress being made here in the Russian Federation and looking forward to being able to look across national boundaries at

issues that really transcend borders, like climate change, gender equality, and education. So I think that we have an exciting future ahead of us.

A. Stott:

Sören?

S. Auer:

I am also very excited to see what is going on here in Russia and it was quite a quick start: so much has happened in just one year. It will take, of course, more time to really build an ecosystem from small companies, start-ups, citizens, citizen initiatives, getting involved, and educational institutions, research institutions, to fill this with life. I look forward to observing that over the coming years.

A. Stott:

OK.

M. Abyzov:

Francis Maude, Minister for the Cabinet Office of the United Kingdom and the man responsible for the open data concept, surprised me at the open data roundtable in London. He said that transparency is tricky and takes time. I believed that this was true in the Russian Federation and that other countries were already in semi-automatic mode. But in London I heard the same thing that I heard in Russia a year ago: it is uncomfortable for government officials, citizens are calling for it, it is not really understood by the population at large, and sometimes it is hard to implement with help from business. All the problems and successes of open data are absolutely international. We can draw upon the experience of the broad international community that is also implementing the open data concept.

It is okay to talk about ideology when the key decisions have not yet been made and there are many forks in the road. But the key decisions on open data were made a long time ago and the Open Data Charter has been signed by the G8. This charter

is a specific tool for use by the entire industry. It includes requirements for information release by default; it addresses the issue of criteria for access to open data (platform format based on open code); requirements for government regulations of open data; and establishment of innovative trends. Using the key principles of this charter, we must develop the concept not only in Russia, but in international cooperation as well.

A. Stott:

Two final thoughts from me. That G8 Charter is, I think, very significant, and one of the things in there is the specific types of data that governments should release that are economically valuable or very valuable for transparency. Any government can release boring data that no one will ever use. The challenge is to release data that is really going to power the economy and help drive the improvement of public services.

On a personal note, I have been involved for the last year with the World Bank here in Russia on open data. This is my fifth trip to Russia in six months, which shows that my multi-entry visa was a superb investment, and I hope to be doing more. I think that there is enormous potential here, and I think that what is in the pipeline will move Russia up that G8 league table very quickly indeed, and I wish them every success. Any help that I can give, I will be happy to do so.

Finally, we have been very squeezed for time, but I hope the audience will thank our distinguished speakers and our discussants in the traditional way. Thank you very much.