

Проблема открытых данных в России

И.В. Бегтин¹

Ключевые слова: открытые данные, качество информации, машиночитаемая форма, государственные сервисы

Open Data in Russia

I. Begtin¹

Key words: open data, information quality, machine-readable form, state services

Особенность присутствия государственных структур в российском сегменте сети Интернет заключается в том, что большинство органов власти предпочитает скрывать информацию от граждан. Например, МВД России не предоставляет информацию о раскрытии преступлений, в частности, не указывается отделение полиции, ведущее следствие. Данные Росстата публикуются таким образом, что требуются существенные усилия, чтобы получить необходимую информацию. Такая ситуация вызывает у граждан ощущение непрозрачности работы государственных механизмов. В то же время государство на всех уровнях своей деятельности само лишает себя возможности использовать механизмы информационного влияния на россиян.

Уникальность Интернета состоит в том, что люди схожих интересов могут находить друг друга, несмотря на расстояние и род деятельности. Общие интересы зачастую формируют круг активных граждан — тех, кому безразлично то, что происходит со страной. Социальная организация, привлечение внимания россиян к деятельности различных ветвей власти, создание среды, в которой каждый может помочь другим, — все это государство может обе-

The peculiarity of state's presence in the Russian segment of the Internet is that the majority of authorities prefer to hold back information rather than disclose it to the citizens. For instance, the Ministry of Internal Affairs of Russia has yet not started to provide information on crime detection with detail at least at the precinct level or data of Rosstat (Federal State Statistics Service) are published in such a way that one has to apply best efforts to retrieve the data from the website in a user-friendly style. This approach is erroneous not only because citizens feel lack of transparency of state mechanisms but also due to the fact that the state at all activities levels deprives itself of the mechanisms of information influence on citizens being the end users of information. Indeed, the unique feature of the Internet is that people having similar interests may find each other despite distances, positions and occupation. Likeness of interests often forms activity domains of citizens, who care what happens to the country. Social organizing, involvement of people in activities of different authority branches, creation of an environment where everybody could make strong contribution and help others, this all is what the state could provide directly by creating Internet-resources helping

¹Генеральный директор ООО «Лаборатория интеллектуального анализа данных», тел.: +7(495)542-87-44, e-mail: ibegtin@gmail.com

¹General Director of ООО "Laboratorya Intellectualnogo Analiza Dannykh" (Data Mining Laboratory), phone: +7(495)542-87-44, e-mail: ibegtin@gmail.com

спечить как прямо, создавая интернет-ресурсы, так и косвенно — через помощь общественным организациям, готовым открывать новые сайты в сети.

Зачем нужна открытость?

Ключевой вопрос, который, вероятно, задает себе каждый человек, работающий в госструктуре: «Зачем нужно открывать информацию?»

Нынешняя ситуация такова, что социальная роль Интернета растет год от года вместе с числом людей, выходящих в сеть, вместе с каждым новым общественным и коммерческим проектом. При этом граждане, и без того относящиеся к государству с недоверием, сталкиваются с тем, что качество интернет-ресурсов, представляющих госструктуры, достаточно низкое. Достаточно сравнить информационные сайты других стран, в частности те, где представлены статистические данные, и сайт Росстата в нашей стране. Сравнение будет не в пользу России. Государственные ресурсы зачастую содержат неполную, устаревшую информацию, представленную в такой форме, что ее дальнейшее использование невозможно.

Наиболее активная часть граждан, безусловно, обращает на это внимание. Недостаточное присутствие государства в Интернете, неполнота предоставляемых данных и отсутствие системного информационного влияния государства в сети отталкивают социально активных граждан от участия в государственной и общественной деятельности. В результате они переходят к общению в политизированных социальных сетях. Таким образом, открытые данные и публичность государственной информации усиливают взаимодействие граждан с государством и обеспечивают поддержку малых информационных компаний (рис. 1).



Рис. 1. Конкурс по открытым данным в России «Apps4Russia»
Fig. 1. Open data competition in Russia “Apps4Russia”

people to self-organize and indirectly by assisting public organizations, which are ready to create such resources with the support of the state.

Why?

The key question everybody working in a government institution or with government institutions is likely to ask himself is: “What’s the need in data openness? Why do we need open data?”

The current situation is that the social role of the Internet becomes bigger year after year along with the number of people having connection to the network, along with every new public and commercial project. At the same time citizens already mistrustful of the state come across the fact the quality of its representation on the net is bad and sometimes even horrible. Those who ever worked with statistics of other countries and came across data of Rosstat in Russia know what I’m talking about. State resources often contain incomplete and outdated information, sometimes due to illiteracy and sometimes due to intentional representation in a form unsuitable for its reuse.

The most active part of citizens (from both political and economical standpoint) can’t stop noticing that. Insufficient presence of the state on the Internet, incompleteness of provided data and lack of a system approach to the informational influence of the state on the net hold socially active citizens back from participation in the state and socially beneficial activities and push them toward participation in politically charged social networks. Therefore, the answer to the question “Why...?” is known. Open data and public nature of state information enforce interaction of citizens with the state and ensure support of small information companies.

How?

Enforcement of presence of the state on the Internet could follow various ways, but the primary way is the departure from the rule “no extra information” towards constant and steady intensification of information influence. Amount, quality and form of disclosed state information shall become high priorities in the state’s presence on the net, in organizing new network projects, portals, web-resources of the state authority bodies.

For instance, to the present day there is no information resource in Russia, which could be called the state’s representative. Despite the fact that there is the website www.gov.ru, it could by no means claim this role both due to actual incompleteness and irrelevance of data posted on the site and minimum friendliness of the provided information. All other state resources cover only information cross-sections within their competence.

Как это возможно?

Усиление присутствия государства в сети Интернет может происходить различными путями, но главный из них — это отказ от установки «никакой лишней информации», что приведет к постоянному и неуклонному возрастанию информационного воздействия на социум. Государство должно заботиться о достаточных объемах, высоком качестве и удобной форме подачи информации в Интернете, об организации новых сетевых проектов, порталов, веб-ресурсов органов государственной власти.

Например, сегодня в России отсутствует информационный ресурс, который можно было бы назвать лицом государства. Несмотря на существование сайта www.gov.ru, он никак не может претендовать на эту роль в связи с неполнотой и неактуальностью размещенных на нем данных, а также по причине неудобной формы размещения данных. Все остальные государственные ресурсы представляют материалы узкой направленности, ограниченные рамками определенной сферы.

России необходимо сформировать «собственное лицо» в Интернете. Важно создать портал, где содержится информация для россиян и иностранцев, желающих узнать о культуре, экономике России.

Технологичность и удобство

В погоне за инновационностью, которая стала проявляться на некоторых государственных ресурсах (появились видеотрансляции, интерактивные сервисы), нередко теряется технологичность их создания.

В то время как государственные учреждения во всем мире размещают информацию в открытом доступе, а такие проекты, как W3C Linked Data позволяют организовывать ресурсы в Интернете как единую информационную базу данных, в России большая часть веб-сайтов до сих пор не имеет RSS лент, т.е. недостаточно удобна в использовании. Многие российские сайты публикуют аналитическую информацию в формах и форматах, непригодных для последующей обработки. В этом Россия разительно отличается от других стран, в частности стран Евросоюза и США, что заставляет задуматься об истинных причинах создавшейся ситуации: либо это непонимание существующей проблемы в информационной сфере, либо осознанные действия государства. Необходимы качественные изменения, в том числе привлечение коммерческих и некоммерческих компаний к формированию и поддержке «поля» открытых данных, что должно вызвать интерес к раскрываемой государством информации у максимально широких слоев граждан.

Правильно организованными источниками данных можно назвать такие проекты, как Data.gov и Data.gov.uk.



Рис. 2. Государственный портал по открытым данным США Data.gov
Рис. 2. USA open data state portal Data.gov

Russia needs to build “its own face” on the Internet. It is important to establish a portal — common entry point where all information would be contained for both residents and non-residents, who want to learn about Russian culture, economy, etc.

Technological Intensity and Friendliness

In pursuit of conventional innovativeness, which has become apparent on some government resources (now featuring video broadcasting and interactive services) technological intensity is often lost in their development.

While the entire world follows the path of disclosing information by governmental institutions, and such projects as W3C Linked Data allow organizing resources on the Internet in the form of a unified information database, the largest part of information resources in Russia are still missing RSS feeds. Their user-friendliness is only fair. They publish statistical and other analytical information in forms and formats unsuitable for any follow-up processing. This is so different from the other countries, particularly the European Union and the USA, that it sets to thinking: “What is it? Misunderstanding or intention?” There is a need in qualitative changes including involvement of commercial and non-commercial companies in forming and maintaining “the field” of open data, which should attract maximum public to information disclosed by the government.

Examples of correctly organized data sources could be such projects as Data.gov and Data.gov.uk. Those government websites of the USA and Great Britain disclose machine-readable information on the state activities. Depending on the type of information it is represented in different styles. Geographic location data are exhibited in KML formats and in specialized formats for GIS-systems, state expenses data are

Через эти правительственные сайты США и Великобритании раскрывается машиночитаемая информация о деятельности государства (рис. 2). Она представлена в разных формах. Данные о географическом местонахождении — в форматах KML и специализированных форматах для ГИС-систем, данные о госрасходах — в форматах CSV и XML, статистические данные — в XML, CSV, SDMX и множестве других форматов. В самых сложных случаях, когда необходим доступ к большим банкам данных, он предоставляется через специализированные веб-сервисы. Причем каждый из источников и массивов информации сопровождается описанием метаданных.

Полная информация. Нулевые затраты

Типовую модель государственной автоматизации в России можно описать так: «Давайте попросим много денег и что-нибудь сделаем». Такой подход к решению задачи приводит к тому, что появляются веб-ресурсы, которые значительно хуже уже существующих, в том числе коммерческих, и они являются результатом несравненно больших затрат. Недавние интернет-проекты подобного рода, такие, как федеральный сайт закупок или портал работы ТруdVсем, являются характерными примерами.

А ведь роль государства и его затраты, особенно в период кризиса, могут быть совершенно иными. Основой изменений может стать переход от разработки невостребованных ресурсов к организации конкурсов по раскрытию информации государственных структур и ее применению. Если государство будет предоставлять материалы в машиночитаемых форматах, то госзатраты могут быть нулевыми, поскольку все коммерческие ресурсы будут заинтересованы в публикации информации, которая может привлечь к ним большее число посетителей. Требования к машиночитаемости информации нигде не закреплены. Из нескольких десятков, если не сотен созданных реестров самого разного толка лидерство по прозрачности уверенно держит только Федеральное казначейство России, раскрывающее информацию о государственных контрактах. Другие госресурсы в лучшем случае предлагают информацию в виде Excel-файлов, а в худшем — в виде отсканированных изображений или файлов для Microsoft Word и PDF-документов.

Вопросы качества

Качество информации, публикуемой государственными органами, — это, несомненно, очень острый вопрос и для тех, кто эту информацию использует, и для тех, кто избегает ее раскрытия. Причиной данной проблемы является низкий уровень информационной культуры работников значительной части российских государственных учреж-

shown in CSV and XML formats, statistical data — in XML, CSV, SDMX and many other formats. In the most complex cases requiring access to large data banks it is granted via specialized web-services. What's more each source and data array is accompanied by metadata description.

Full Information. Zero Expenses

A typical model of state automatization in Russia can be described as follows: “Let's ask for much money and do something”. Such approach to tackling an objective leads to appearance of web-resources, which are way worse than those that already exist including those of commercial nature, and they are built at incomparably greater expenses. Recent resources of the kind, such as the federal procurement website or the job opportunities portal TrudVsem can be exemplary to that.

After all, the role of the state and its expenses especially in the crisis period could be absolutely different. Changes could be based on transition from development of uncalled resources to organization of tenders for disclosure of information provided by the state and for its application. If the state discloses information in machine-readable formats, state expenses could be zero, as all commercial resources would be interested in publishing information that could attract more visitors. Requirements to machine-readability of information disclosure are not defined. Out of several tens, if not hundreds of built registers of different kinds, strong leadership belongs only to the Federal Treasury, which discloses information on state contracts. Other state resources offer information in the form of Excel spreadsheets at best or in the form of scanned images or files for Microsoft Word and PDF at worst.

Quality Issues

Quality of information published by state agencies is indeed a very sensitive issue for those who use this information and for those who avoids its disclosure. The root cause of this problem is the low level of informational culture of employees of the most Russian state institutions and lack of procedures for collection, processing and updating of developed data arrays. This leads to the fact that in open data discussions state officials try to refuse from their disclosure explaining it by their low quality. But this is exactly information disclosure that can improve the quality!

Here's a telling example — the Clearspending Project of Sunlight Labs, which compared data of state budgets in USA with contract data and found numerous mistakes and discrepancies. Due to this fact, USA's competent agencies are now working on quality improvement of provided information and the report of Sunlight Labs was reviewed by the Congress.

GEOINFORMATION SERVICES
ONE unique portfolio
ONE hundred percent services
GEO-Information just got easier!

Astrium Services' GEO-Information division brings together the expertise, capabilities, and unmatched archives of Spot Image and Infoterra. Whether working in land management or on increasing our security, protecting the environment or monitoring natural resources, decision-makers across the world can now rely on ONE unique point of contact for access to innovative and highly-competitive end-to-end solutions. With 15,000 employees and a turnover of €5 billion, Astrium is a global leader in the domains of Space Transportation, Satellites and Services. contact@astrium-geo.com

ASTRIUM
 AN EADS COMPANY

дений и отсутствие регламента сбора, обработки и обновления создаваемых ими материалов. Чиновники пытаются отказаться от раскрытия данных, ссылаясь на их низкое качество. Но именно раскрытие информации и способно это качество повысить!

Наглядный пример — проект Clearspending от Sunlight Labs, который сопоставил данные госбюджетов США с данными об американских контрактах. В результате были обнаружены множественные расхождения и ошибки. Отчет Sunlight Labs рассматривался в конгрессе. Благодаря этому ответственные структуры США сейчас занимаются повышением качества предоставляемой информации. Иначе говоря, если государственные структуры хотят решить эту проблему, они **ДОЛЖНЫ** публиковать информацию, а не скрывать ее.

Обратная сторона медали — огромное количество открытых данных, которые представлены в форматах, непригодных для машиночитаемой обработки. Это PDF, Word, HTML-документы, реестры различных объектов и т.д. Например, ежедневные отчеты Мосгорздрава можно преобразовать в открытые данные, некоторые из них уже опубликованы. Однако им необходимо придать удобную машиночитаемую форму.

In other words if state authorities want to resolve the issue about raw data — they **SHALL** publish them but not hold back.

On the other hand there is enormous amount of data, which are already published but in formats unsuitable for machine-readable processing. These are PDF, Word, HTML documents, registers of various facilities, etc. For example, there are daily reports of Moscow Health Department, which can be converted into open data and which are already open. These are not raw data! They are already published! They only need to be shaped in a friendly machine-readable form.

Signs of Reforms

However many changes are under way. Lately Russian officials have finally focused on the issue of open data, more often words in support of open data are heard from representatives of the Ministry of Economic Development of Russia and the Administration of the Moscow City. Officials have finally come to understanding that involvement of public society and business in reusing state information enhances confidence of citizens in activities of state agencies, supports civil initiatives, and helps the existing Internet-companies and first-time entrepreneurs to improve performance.

Преобразования

Сейчас многое меняется. В последнее время чиновники в России наконец-то обратили внимание на проблему открытых данных, всё чаще слова в поддержку «рассекречивания» информации звучат из уст представителей Минэкономразвития России и администрации города Москвы. Чиновники пришли к пониманию того, что вовлечение общества и бизнеса в повторное использование государственной информации повышает уровень доверия граждан к деятельности госструктур, поддерживает гражданские инициативы, помогает существующим интернет-компаниям и начинающим предпринимателям эффективнее работать.

Безусловно, многие не уверены в том, что представители власти будут раскрывать действительно важную информацию, что их усилия будут последовательными и долгосрочными. Некоторые полагают, что эта идея погибнет в бюрократической машине. Все эти риски существуют, однако мы не должны бездействовать. Сегодня проблема открытых данных является ключевой темой во всех развитых странах. Это во многом заслуга тех граждан и организаций, которые использовали свою волю и навыки, чтобы добиться цели.

Наша роль как россиян, как предпринимателей, как потенциальных пользователей этой информации заключается в том, чтобы выразить свое мнение и сообщить его представителям власти. Помимо того, что мы можем поднимать вопрос об открытости данных при каждой встрече с ответственными чиновниками, нам необходимо внести самостоятельный вклад в создание качественных материалов. Существует немало инструментов по преобразованию полуструктурированной информации в машиночитаемую форму, приданию этим данным описания, их интеграции в существующие коммерческие и некоммерческие проекты. Мы способны все это сделать в соответствии с последними мировыми тенденциями.

Indeed, many of us remain skeptical as to whether officials are ready to disclose really important information and whether their efforts will be consistent and long-term and whether this idea will survive the bureaucratic machine. All these risks exist however we shall not sit idly. Nothing but our efforts can overcome them and the fact that open data have become the key topic of discussions in all developed countries is in the first place due to those citizens and organizations that exercised their willpower and skills to achieve this objective. Our role as citizens, entrepreneurs, and potential users of this information, is to drive our opinion to the authorities' representatives. In addition we can raise the issue on data openness at any meeting with competent officials and we also can make our own contribution to their development. There are a lot of tools for converting semi-structured information in machine-readable form, for giving descriptions for those data, for their integration into existing commercial and non-commercial projects. We are capable of doing that and this is exactly how things are going around the world.