

# Development of the St. Petersburg's Linked Open Data Site Using Information Workbench

Dmitry Mouromtsev, Vitaly Vlasov, Olga Parkhimovich, Mikhail Galkin, Vitaly Knyazev  
 National Research University of Information Technologies, Mechanics and Optics  
 Saint Petersburg, Russia  
 dimour@mail.ru, {inxao, olya.parkhimovich, mikhgalkin}@gmail.com, gg.nexus@hotmail.com

**Abstract**—This paper discusses the Russian projects publishing open government data. The article also describes the development of the open linked data portal and its approach to convert open government data in the open linked data. Information Workbench is used to build this system. It allows storing, visualizing and converting data files in Semantic Web formats.

**Keywords**—Government Data, Information Workbench, Ontology, Open Data, Open Government, Semantic Web.

## I. INTRODUCTION

There is a trend nowadays to publish open government data, such as public digital information in the Internet that can be analyzed or used. Due to this fact lots of citizens' problems can be solved by means of different services or projects, large possibilities for commercial activities that can improve government economics arise, government departments can improve their efficiency and the governance itself becomes more transparent.

During the analysis of Russian and world experience, we can consider several projects using open data. For instance, there is a site in the USA [1] that helps citizens to find out information about country's budget funds' forming and spending, and also their influence on voting or government policies. Similar service in Australia [2] gives you the opportunity to compare indicators of different regions (level of education, social-economic opportunities, and economic development) and My School portal [3], combining information of all government and private schools: number of students, average marks or financing. The data can be represented in tables, graphics or graphs.

The appearance of the Open Government Partnership [4] in 2011 declares intentions of countries in open data sphere development, transparency growth and governments accountability. Such initiative unites countries that undertake an obligation of increasing transparency.

## II. THE FOREIGN OPEN DATA SITES

Open data appeared in 2009 and still remains a hot topic. The USA is one of the first countries in which the state launched an open data website [5]. Its launch was held in May 2009, and its aim was to improve access to data in

machine-readable formats. At the time of publication of the portal has been loaded 46 datasets.

To obtain feedback from citizens the government has launched a platform in order to collect useful ideas. As a result, 891 people left 175 ideas to refine the portal.

A number of mechanisms has been developed to publish data sets easily and comprehensively:

- Dataset Management System (closed part of the portal).
- API (interface for data publication).
- Metadata Feed (a special service similar to RSS or Atom).

On the other hand, the government has also developed criteria for evaluating the performance of this portal:

- A number of loaded raw datasets.
- Mechanisms for data mapping.
- Statistics of datasets downloads.
- Statistics of visits of this portal [6].

U.S. open data portal allows you to download data files, display them and visualize. Also, it contains semantic web technologies and applications that are based on open data (Fig. 1).



Fig. 1. U.S. official open data site.

Publication of this data in the USA is supported by the law and governed by the following major legislative acts: the White House Open Government Initiative [7], the President’s Memorandum on Transparency and Open Government [8] and the Open Government Directive [9].

Another leader in the field of open data is the United Kingdom, which has launched the open data portal [10] in January 2010. The number of data sets at the time of publication was nearly 2500. The main part of those data sets was topographic service and state statistics [11]. Publication of open data in the UK is also regulated by the law. Key government acts in this area are the Open Data White Paper. Unleashing the Potential [12], the Protection of Freedoms Act 2012 [13], The Power of Information: An independent review by Ed Mayo and Tom Steinberg [14].

The main functional blocks on the UK open data portal are described below (Fig. 2):

- The data (search for tags, departments, locations, cost of the publication).
- Community (blogs, ideas, forums).
- Requests for data publication (forms and dashboards).
- Applications (the gallery application, recently added applications and their assessment, rss-subscription of new applications).
- Library (filter by category, tags, sector and type of documents).



Fig. 2. The UK’s official open data site.

Open data portals also exist in other countries, such as Canada [15]. Its website has been launched in the spring of 2011 and had been testing for almost a year. When the portal has been launched, there has been published over 800 datasets, but a year later the number reached 11,000. Today, the portal contains 197 831 datasets published by 25 agencies.

After launching the site the “Canada’s Action Plan on Open Government” [16] has been submitted. Nowadays the document is the basis related to the area of transparency and open data.

The main functional blocks in this site are (Fig. 3):

- Data Search (top 10 data sets, filters for agencies, a type, tags, a format).
- Open data (best practices, an interactive map with links to the sites of open data territories and municipalities, the facts and figures, developers).
- Open government (where citizens give their feedback on presented data, quarterly financial reports, future publications).
- About the portal (open license, the principles of publication, frequently asked questions).
- Participation (offer a data set, a blog page developer, consulting, application) [17].



Fig. 3. The Canada’s official open data site

### III. THE HISTORY OF OPEN DATA AND OPEN GOVERNMENT IN RUSSIA

For a long time, the main problem of the development of open data in Russia was the absence of legislation and generally accepted terminology in this field. As a result, different agencies used different concepts for the term "open data", thus making it difficult to understand the importance of open data.

In April 2012, the Russian Ministry of Foreign Affairs sent a letter to the Open Government Partnership, which declared intentions to join the partnership. After that the Russia's Civil initiatives platform for open government was established [18]. But on May 17, 2013 Vladimir Putin announced the stop of Russia's accession to the OGP [19].

September 2, 2011 at the meeting of the Presidium of the Presidential Council for the Development of Information Society a number of proposals was made in order to establish the platform to publish open data [20].

In October 2011, Dmitry Medvedev took the initiative to development as an expert platform "Big Government" [21]. It has become a tool of interaction between the

government and the expert community, corresponding to the ideas of open government.

February 8, 2012 by the decree of the President of the Russian Federation a working group was formed to prepare proposals for the formation of "Open Government". After that the crucial goals of open government were announced:

- Fighting corruption.
- Development of competition and entrepreneurship.
- Development of human resources and human capital [22].

Resolution of the Government of Russia on July 26, 2012 can be found in [23].

Then, the document supervising development of Open Government was approved on 26 September, 2012 [24].

#### IV. RUSSIAN OPEN DATA PROJECTS

Since 2013 open data portals appear in different regions of the Russian Federation.

The first open data portal within the site of the Moscow's open data [25] has been launched in January 2013 by the Moscow government. It contains datasets in various readable formats. Initially, more than 150 datasets have been uploaded, for example:

- Address youth centers.
- Objects overhaul of the road network.
- The exact boundaries of regions.

On the Moscow's open data portal you can find the following functional blocks (Fig. 4):

- List of loaded datasets.
- List of applications based on Moscow's datasets.
- Moscow map (with the possibility to view different layers).
- Information for developers.



Fig. 4. The Moscow's official open data site.

To increase the number of users who use the data, government run contests for developers of applications and announce to the media about the new datasets.

After a few months regional governments have begun to develop their open data portals. For example, in June 2013 the open data portal of the Ulyanovsk region has been launched [26]. It might be considered as an important step to the implementation of the Initiative about data publication. During the launch of the portal only one data set has been uploaded, but now the number has increased to 43.

At the moment, the entire data is published in machine-readable formats and is available in two versions: in CSV format and in tables (Fig. 5). This allows one to browse a particular data and refine it.

The Government of the Ulyanovsk region cooperates with the World Bank in the field of open data. In October 2012 the Memorandum of Understanding was signed between them. The development of the open data portal is based on the principles described in the document "Evaluation of ready to use open data", which was carried out by the World Bank.

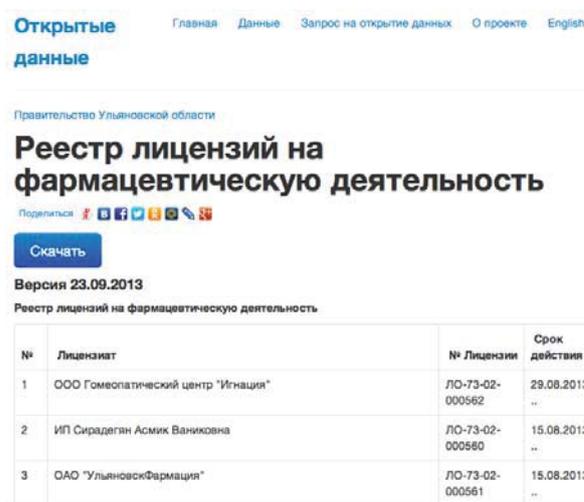


Fig. 5. The Ulyanovsk region's official open data site.

Currently Russia's federal open data portal is being developed. But some departments and agencies are beginning to publish open data on a separate page on their websites.

For example, Fig. 6 shows an example of placing an array of open data from the Ministry of Education and Science of the Russian Federation.

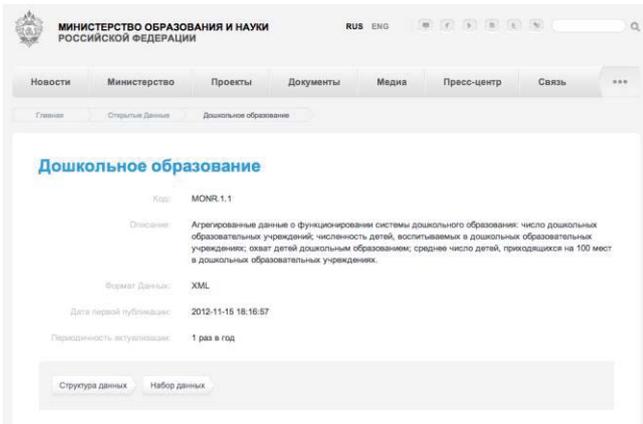


Fig. 6. Page of open data on the portal of the Ministry of Education and Science of the Russian Federation

Similar pages can be found within websites of the Ministry of Interior [27], Ministry of Justice [28], Ministry of Economic Development [29]. In some cases, government agencies create special sites to place open data, but most often information is placed on a separate page of the website of the relevant department.

Attempts to publish open data by states are a big step towards transparency. But in case of unification and machine processing, the data should be published in linked data formats. For example, the recommendations by W3C [30] or the guidelines developed by Socrata (a platform to build sites for publishing open data) [31].

So we started to develop a system that would allow automatic conversion the data to semantic web formats and its further publishing.

V. TECHNICAL ASPECTS

According to the theory described above, appropriate realization and architecture of the system must be developed. The architecture of the system should be designed under certain requirements and possess certain features:

- The ability to gather open data from different sources.
- The ability to transform this data to RDF format; Provide SPARQL endpoint in order to access open data in the system.
- Present data in a number of formats, e.g. wiki, tables, charts, graphs.

Therefore, an appropriate technical platform is needed. There are several solutions which could be used as a platform:

- Semantic MediaWiki (SMW).
- Information Workbench (IWB).
- OntoWiki.

Each system has its own advantages and disadvantages, so it's better to use a table in order to compare them (Table I).

TABLE I COMPARISON OF SEMANTIC PLATFORMS.

	SMW	IWB	OntoWiki
Programming language	PHP	Java	PHP
Developer	Institute AFIB at Karlsruhe Institute of Technology (KIT)	fluid Operations	AKSW research group, Leipzig
Dependency	MediaWiki extension	Standalone	Standalone
Database	4Store Virtuoso	Sesame	MySQL Virtuoso
Query Language	SMW-QL SPARQL	SPARQL	SQL SPARQL
Data Visualizing	Rich	Rich	Rich

We believe it advisable to use The Information Workbench (IWB) by FluidOps. It has an interesting concept of Data Providers, therefore it might be the best for us. With the use of this mechanism it becomes possible to develop an extension that can update automatically uploaded data from a particular source. Also it's a beneficial advancement that IWB is written in Java.

The Information Workbench in itself is a powerful tool to build systems working with all kinds of semantic data. IWB satisfies all the requirements and provides additional advancements and features, which are connected with methods of parsing data, representing data and accessing data:

- Works with a variety of formats (RDF, N3, Turtle, N-Triples, TriG, TriX).
- Provides broad opportunities to visualize data (charts, tables, graphs, maps, etc.).
- Has a built-in SPARQL endpoint.
- Supports the concept of Linked Open Data and provides tools to work with this data.
- Has a detailed documentation and comprehensive API.

The platform possesses a lot of ways of representing data. Some of them are shown below:

- Tables (Fig. 7).
- Charts (Fig. 8).
- Graphs (Fig. 9).
- Maps (Fig. 10).

Due to flexible architecture of the chosen platform, the system based on it possesses a strong potential to future improvements and adjustments.

Code	Name
AD	Austrian Development Bank
AE	Austrian Development Bank
AF	Austrian Development Bank
AG	Austrian Development Bank
AI	Austrian Development Bank
AL	Albanian Development Bank
AM	Armenian Development Bank
AN	Andean Development Bank
AO	Angolan Development Bank
AP	Andean Development Bank
AR	Argentine Development Bank
AS	Asian Development Bank
AT	Austrian Development Bank
AW	Austrian Development Bank
AX	Austrian Development Bank
AY	Austrian Development Bank
AZ	Azerbaijani Development Bank
BA	Bosnian Development Bank
BB	Bahian Development Bank
BC	Bahian Development Bank
BD	Bangladesh Development Bank
BE	Belgian Development Bank
BF	Burkinabe Development Bank
BG	Bulgarian Development Bank
BH	Bahraini Development Bank
BI	Burundian Development Bank
BJ	Burkinabe Development Bank
BK	Burkinabe Development Bank
BL	Burkinabe Development Bank
BM	Bermudian Development Bank
BN	Bruneian Development Bank
BO	Bolivian Development Bank
BP	Bolivian Development Bank
BQ	Bolivian Development Bank
BR	Brazilian Development Bank
BS	Bahian Development Bank
BT	Bhutanese Development Bank
BV	Brazilian Development Bank
BW	Botswana Development Bank
BX	Brazilian Development Bank
BY	Belarusian Development Bank
BZ	Belizean Development Bank
CA	Canadian Development Bank
CB	Canadian Development Bank
CC	Caribbean Development Bank
CD	Congolese Development Bank
CE	Caribbean Development Bank
CF	Congolese Development Bank
CG	Congolese Development Bank
CH	Swiss Development Bank
CI	Cote d'Ivoire Development Bank
CJ	Caribbean Development Bank
CK	Caribbean Development Bank
CL	Chilean Development Bank
CM	Congolese Development Bank
CN	Chinese Development Bank
CO	Colombian Development Bank
CP	Caribbean Development Bank
CQ	Caribbean Development Bank
CR	Costa Rican Development Bank
CS	Czech Development Bank
CT	Caribbean Development Bank
CU	Cuban Development Bank
CV	Cape Verdean Development Bank
CW	Curaçaoan Development Bank
CX	Caribbean Development Bank
CY	Cypriot Development Bank
CZ	Czech Development Bank
DA	Danish Development Bank
DB	Dominican Development Bank
DC	Dominican Development Bank
DD	Dominican Development Bank
DE	German Development Bank
DF	Dominican Development Bank
DG	Dominican Development Bank
DH	Dominican Development Bank
DI	Dominican Development Bank
DJ	Dominican Development Bank
DK	Danish Development Bank
DL	Dominican Development Bank
DM	Dominican Development Bank
DN	Dominican Development Bank
DO	Dominican Development Bank
DP	Dominican Development Bank
DQ	Dominican Development Bank
DR	Dominican Development Bank
DS	Dominican Development Bank
DT	Dominican Development Bank
DU	Dominican Development Bank
DV	Dominican Development Bank
DW	Dominican Development Bank
DX	Dominican Development Bank
DY	Dominican Development Bank
DZ	Dominican Development Bank

Fig. 7. Method of visualizing example data – a table.

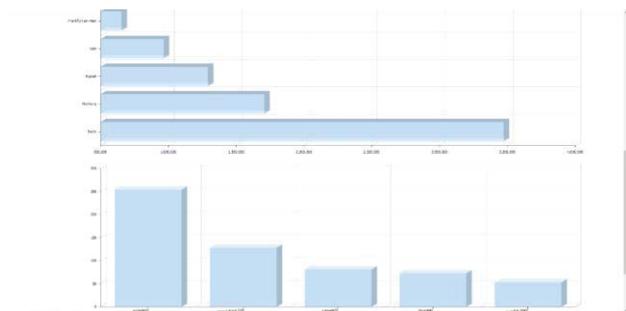


Fig. 8. Method of visualizing example data – a chart.

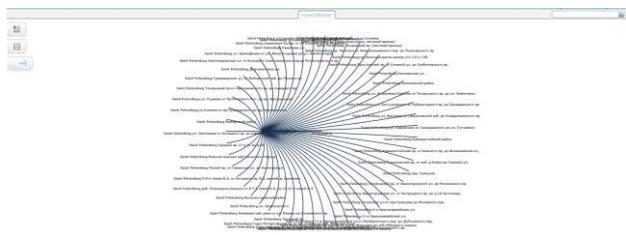


Fig. 9. Method of visualizing example data – a graph.



Fig. 10. Method of visualizing example data – a map.

Current dataset (partially depicted above, fig. 9, fig. 10) is the data that is connected with roads in Saint Petersburg. The data is divided by districts and contains information about lengths of the roads and sidewalks to be repaired. Source table is provided by the official web site of the government of Saint Petersburg. Using presented data it's possible to attach additional statistics, such as traffic jams, parking information, a number of injuries during the year or an average cost of a particular road.

Our initial goal is to develop an extension, which allows user to acquire particular data from HTML pages. It might seem as an easy problem, but current methods of front-end development give a tremendously high number of possible

variants where required data is stored within an HTML code. Hence, we need an adaptive algorithm that is able to search a particular data on every given page. A perfect adaptive algorithm will need only a URL address and automatically decide what data should be collected from a page. A simple algorithm needs at least a title of required data so it can navigate through a page and search for it.

This feature can be implemented using the Data Provider mechanism presented in IWB. Data providers are configurable programs that connect to remote data sources (such as files, Web endpoints, or legacy systems), extract data from these sources, convert it into the RDF data format (where necessary), and store the extracted data in the Information Workbench's local repository.

The Data Provider extension allows one to specify a particular URL with data (for instance, a table) and get this information in RDF format in repository. The algorithm that solves this problem is described below:

1. Determine the structure of the given page.
2. Determine a method to be applied to gather necessary information.
3. Localize data within HTML code (for the example above – localize and get some content of <table> tag).
4. Parse acquired data in RDF format using (initially) predefined ontology.
5. Put the data into a repository.
6. Represent the content in a formal way (e.g. in graphics, in charts, in text).

The algorithm in its nutshell can be depicted on the Fig. 11.

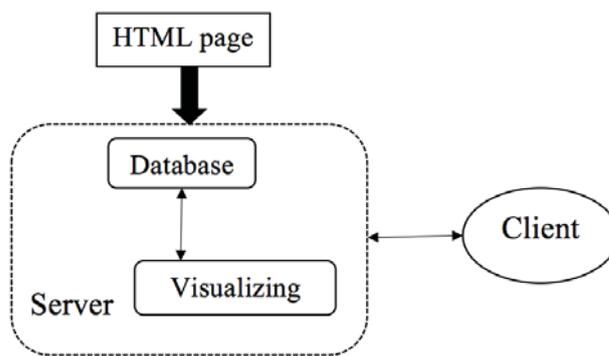


Fig. 11. Extension's architecture

VII. CONCLUSION

In conclusion, it should be noted that open data is an important and urgent topic. Russian projects presented in this paper show that the developments in this area still are not perfect and require significant improvements.

It is also necessary to improve the format in which you publish open government data. Integration of semantic data is an important goal. The Information Workbench is a

powerful tool to help in achieving this goal. The suggested method of unifying semantic data is flexible and beneficial, so it can be upgraded and used for achieving higher goals and objectives.

REFERENCES

- [1] Center for responsive politics official website, Web: <http://www.opensecrets.org>.
- [2] Australian suburban trends official website, Web: <http://www.suburbantrends.com.au/>.
- [3] My school official website, Web: <http://myschool.edu.au/>.
- [4] Open government partnership official website, Web: <http://www.opengovpartnership.org/>.
- [5] An official web site of the United States Open Data, Web: <http://www.data.gov>
- [6] Blog about government sites and government services in the Internet, How created the portals of open data in foreign countries – the United States, Web: <http://gov-gov.ru/?p=3367>
- [7] White House official website, Open government initiative, Web: <http://www.whitehouse.gov/open>
- [8] White House official website, Memorandum for the heads of executive departments and agencies. Subject: Transparency and open government, Web: [http://www.whitehouse.gov/the\\_press\\_office/TransparencyandOpenGovernment/](http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/)
- [9] White House official website, P.R. Orszag “Memorandum for the heads of executive departments and agencies, Subject: Open government directive”, M10-06, Dec. 8, 2009, Web: <http://www.whitehouse.gov/open/documents/open-government-directive>
- [10] The official site of UK opening up government, Web: <http://data.gov.uk>
- [11] Blog about government sites and government services in the Internet, How created the portals of open data in foreign countries – the United Kingdom, Web: <http://gov-gov.ru/?p=3400>
- [12] Open Data White Paper. Unleashing the Potential, Web: [http://data.gov.uk/sites/default/files/Open\\_data\\_White\\_Paper.pdf](http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf)
- [13] Protection of Freedoms Act 2012, Web: <http://www.legislation.gov.uk/ukpga/2012/9/part/6/enacted>
- [14] The Power of Information: An independent review by Ed Mayo and Tom Steinberg, Web: <http://www.opsi.gov.uk/advice/poi/power-of-information-review.pdf>
- [15] The official website of open data in Canada, Web: <http://data.gc.ca/>
- [16] The official website of open data in Canada, “Canada’s action plan on open government”, Web: <http://data.gc.ca/eng/canadas-action-plan-open-government>
- [17] Blog about government sites and government services in the Internet, How created the portals of open data in foreign countries – the Canada, Web: <http://gov-gov.ru/?p=3463>
- [18] Official website of project “Civil initiatives for open government”, Web: <http://www.ogp-russia.ru/>
- [19] M. Ivanov “The Government of Russia to work more discreetly”, Web: <http://www.kommersant.ru/doc/2192570>
- [20] The decision of the Council meeting of the Presidential Council for the Development of Information Society in Russia on Sep. 2, 2011, Web: <http://ictgov.ru/news/1248/>
- [21] Official website of Russian open government, Web: <http://xn--80abeamcuufxbhgound0h9cl.xn--p1ai/>
- [22] Decree of the President of the Russian Federation dated Feb. 8, 2012 No 150 “On the formation of a working group of the Open Government in the Russian Federation”
- [23] The decision of 26 July 2012 No 774 "On the Expert Council under the Government of the Russian Federation"
- [24] Executive Order dated by September 26, 2012 No 1782-r "On the composition of the Government Commission for the coordination of open government"
- [25] Official website of Open data portal of Moscow’s government, Web: <http://data.mos.ru>
- [26] Official website of Ulyanovsk’s open data, Web: <http://data.ulgov.ru>
- [27] Official website of Ministry of Interior, Open data page, Web: <http://mvd.ru/opendata>
- [28] Official website of Ministry of Justice, Open data page, Web: <http://minjust.ru/ru/opendata>
- [29] Official website of Ministry of Economic Development, Open data page, Web: <http://www.economy.gov.ru/wps/wcm/connect/economylib4/designElements/opendata>
- [30] Publishing Open Government Data. W3C Working Draft 8 Sep.2009, Web: <http://www.w3.org/TR/gov-data/>
- [31] 3 Guidelines for Publishing Your First Open Data Sets, Web: <http://www.socrata.com/blog/3-guidelines-for-publishing-your-first-open-data-sets/>