ANALYSIS PLATFORM FOR THE PRESENTATION OF A SET OF OPEN DATA IN EDUCATION AND PUBLIC ADMINISTRATION

Srdjan Atanasijević [1], Milos Miladinovic [1], Vladimir Nedic [2], Milan Matijevic [3]
[1] *ComTrade Solution Engineering, Belgrade, srdjan.atanasijevic@comtrade.com, milos.miladinovic@comtrade.com*
[2] *FILUM (Department of Philology & Arts), Kragujevac, vnedic@kg.ac.rs*
[3] *Department of Engineering Science, Kragujevac, matijevic@kg.ac.rs*

Summary: *Open data is data that can be freely used, processed and exchanged by all users. Data resulting from the processing and by changing the open data applies the same principle: they can continue to be used freely. In this paper we show the application of the basic concepts of open data: standardized models of open data, the process of opening the database, application areas and experience of using the platform for the presentation of open data. In the second part we will make an analysis of the most popular platforms for presentation open data and provide guidelines for their use in our practice. In conclusion, explained the potential application of a set of Open Data in Education and the Serbian economy.*
Key words*: open data, linked data sets, CKAN, opened the public sector*

1. CONCEPT OF OPEN DATA

The concept of open data, open data and more accurately, which are owned by the public sector (state, region, local government, science) it is relatively new. First experience intense opening data dating back to 2009 by the U.S. administration, the United Kingdom, New Zeeland, Canada. The basic idea was to answer the question: where and how to spend the taxpayers' money?
Today, after five years, the concept is evolved became widely accepted and experienced full recognition [1, 2]. Developed IT standards and platforms that facilitate the opening of the data according to the principles that will be exhibited in the OVM work.
What we consider open data which is their added value, how to use them?
Open data is data that can be freely used, processed and exchanged by all users. For data arising from a change of open data applies the same principle, they can still be used freely [3].
Features that adorn the degree of openness of data are crucial for their further application. The information that we believe should be open to meet the following principles:

- **availability of data and access to it: the data must be available in their entirety,** without any special three-to's for their further use (that are open formats, not related to a software license)
- **reuse of data** and further exchange: the right to use of open data given by the owner must provide further exchange, including processing and linking with other data sets. The data should be in a form suitable for computer processing.
- **universal application:** Anyone interested must be able to access, modify and redistribute the data resulting from the processing of open data. There should be no discrimination towards projects that will continue to use open database, as well as to the group of the data used. You should not be restrictions on the type: not used for business purposes or for use only in the educational process.

Applying a set of open data every day is increasing. Figure 1 shows the major field of use and the type of data still open, which include:

- Geo data (GIS, etc.), which are used to plot the map and display objects that are geographically referenced in the application type maps, routes, search by location
- culture, information on events and facts (cultural sites facilities), references (collection data collection)
- science data from scientific experiments, research
- finance, public finance, the data on the expenditure of money from the budget, projects
- statistics, which are the result of the statistical offices, either the Republican or regional
- weather forecast data from the meteorological station, predict the weather, the analysis
- environment, data on the degree of contamination of a site (gases, particles, ..)
- education, public information on the facilities, investment in education projects
- transport, information such as timetables of public transport and so on.

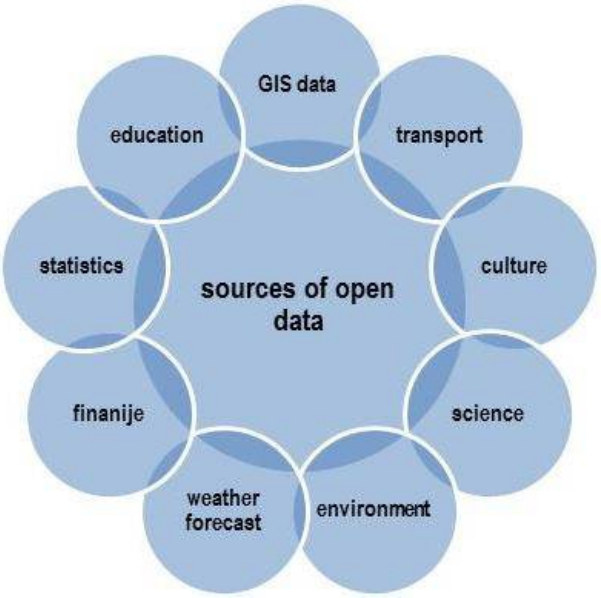

Figure 1 Areas of use set of open data, according to sources

How to turn public data in open data set?

Experience shows that it is to open the data which are owned by the public sector especially important commitment and management decisions, and the implementation of the decision itself and the technical implementation of well-defined in practice. The document, titled *Manual for a description of open data* [3] discusses the legal, social and technical aspects of the opening of public data. The remainder of this paper will present the basic idea of the process of opening the data and we will make an analysis of the most popular platform for the presentation of open data and provide guidelines for their use in our practice [4].

In the **process of** implementation of **the** decision to **open the data** recommended adherence to three fundamental principles:

- make small, simple and effective steps. I can not and should not be open all the data at once. Speed is important as well as the benefit of the data to be opened. Of course, the more open data - to greater effect.
- include as many users as early as possible. Inclusion of data is essential. They should be involved in the selection of data to be open, scope and structure. Users are the engine of the further process. Especially business users to put data into the function of increasing the volume of business and increased revenues will eventually pass through taxes An increase in the public finances.
- principle of openness in communication: to identify key issues and work together to resolve them [5]. Legal, cultural, technical barriers, would, if it does not create a climate of open communication, will lead to the blocking process.

In the **implementation** process, to make ***the*** following ***four steps:***

1 Choosing a dataset that is a candidate for the opening

2 Allow unlimited use of:

- Check whether there are legal obstacles to opening data (Safety or proprietary rights)
- A statement to the data, the legal form which data is safe, and following the principles of open data
- If a candidate for the opening does not meet any of the restrictions, return to step 1

3 put the data available to users. In raw form and / or one of the open form.

4 organization of data in the form of catalogs, with a description of the structure and other information that will make them more applicable in practice users.

2. PRINCIPLES FOR DEFINING THE FUNCTIONS OF THE PORTAL OPEN DATA

To answer the question that all features should have IT systems to support the publication and presentation of open data, the need to resolve the following dilemma:

- how you plan to use open data
- used by the plan to have the use of open data

Ways to access data from a set of open data

In the process of linking open data sets recognize the following characteristic processes:

- quest for the facts from a set of open data
- carrying out useful information from the data set
- display data defining a schema, interfaces and data structures
- access to data via software interface
- data access using the service

1 search for facts Users can search data in search of specific facts. Facts can be found through the available user display or Excel tables. Data → Is the fact Search Performance	**2 from data to information** Creating a static representation and interpretation of one or more data sets. It is commonly used in the visual presentation, review and reporting. Data → Information Management Static analysis Visualization Reports
3 of data to the interface Interactive access and search one or more sets of data. Defines the schema that describes the structure of data and it is an interface for data usage. Data → Interface Combining data sets Writing the code Providing interface	**4 from data to information** Availability of data sets in a variety of formats. Combining multiple sets. The assemblies created by the API, and is defined, so that the data can be accessed and program. Also available is the interface through which data can be downloaded for free. Data → Data Conversion Filtering Combining Providing API Datasets are ready for download
5 the data across services The data is stored behind a set of services that access them. The users of the various services available that are based on open data.	Data → Service Integration with existing services Creating new services

Table 1 Typical methods of using data from a set of open data

These processes are not mutually exclusive, and is the most common case, a combination of several processes. The current situation is such that there are problems when it comes to mapping data and life cycle data sets to the final stage, they become visible to end users. Such problems is less, because the description of the principles of open data continuously improve and enhance [6, 7].

Usability data from a set of open data

In order to utilize the full value of a set of open data that we want to use, it is important that the information and data are placed in a context that will create new knowledge and to enable the creation of new services, applications and services.

Related Information *("linked data")* facilitate the creation of different types of services and applications that are based on interrelated data [6, 7]. Considering the usefulness of open data principles, we begin with two equal point of view with related data, such as the view from the institution that should enable the publication and by the user that in our processes and solutions consuming open data.

Assessment of the usability of open data can be viewed along two main directions (Figure 2):

- the degree of complexity of data access (technical dimension of the problem) and
- the degree of openness / transparency of the content of the data (expressed in the details of the data)

The first direction is technically applicable to the accessibility of the data contained in the file format. The information listed in the PDF files, for example, are less available than data published in structured tables Excel. File formats that are related to the licensed application solutions that interpret them are less available than open source file formats such as CSV.

The clearest gradation scored the opening and connecting the data presented by Tim Berners-Lee [6], presenting his model with five stars:

***The degree** of openness*	Usability of open data: *The user can (details)*	Investment data to be open: *Owners should provide*
*	• review • print • preserve local • manually switch to another system • modified form optionally • share them with others without restrictions	• simply publishing • do not have to explain that the data can be used freely
**	• further processing in the wet with the protection of software: aggregation, visualization, calculation of the any data ... • export to other structured format	• simply publishing
***	• format data is free: it does not require you to pay a license to read the data	• must have modules for commercial export from the free format • simply publishing
****	• not link to any site • save • use of the multiple times • safely combined with other data	• access control and display parts of the data • owners of different data sets can be interconnected • to define templates of their data • you need to invest time in organizing data for display
*****	• to reveal the interconnection of data • to learn about the data directly from the data schemas	• data should have a schema description • increased the value of data • a profit from the publication and use of data as well as user
Legend:		
* Information and data are available on the Web (in any format) under an open license ** The information and data available as structured data (Excel format instead of a scanned image of the table) *** Unprotected formats (non-proprietary) - CSV instead of Excel **** URI identification, so that people can point to individual data ***** Providing context linking different data sets		

Table 2 Model 5 star: the path of the open linked data

Linked data *("linked data")* refers to a set of data that describes an open standard for metadata (such as the Uniform Resource Identifier (URI) and the Resource Description Framework (RDF). Such data described are easily readable for humans and for machines [6].

Another direction considering the open data refers to the types of sources and details of the published data. The most natural source of information in the public sector is the traditional statistics, such as the results of the census of households. Geospatial data is also one of the most commonly published public data and are useful for the majority of applications. Public procurement and data on budgetary spending less transparent data source.

Linked data obtained an increasing role in the field of data management. Are independent of the domain and the areas in which they are used, thus proving its advantages over the traditional method of data management. Format of data is marked with 5 stars is called the associated data.

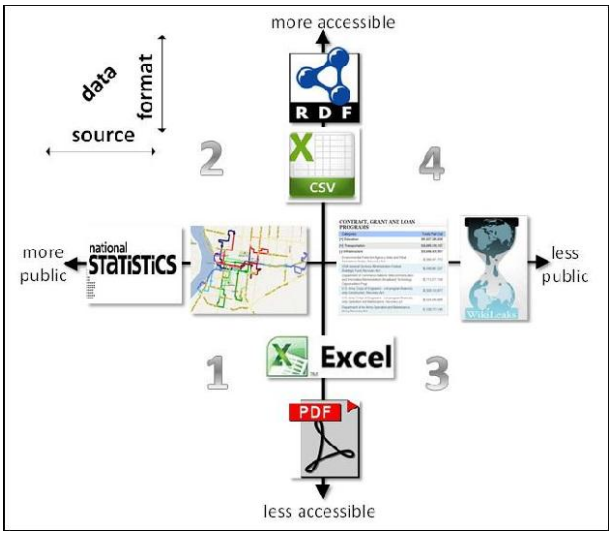

Figure 2 The degree of autonomy of use of open data: according to the file format and contents

Linked data using semantic web standards such as RDF standard to give a description of the data and to connect with other data sets and thus provide the appropriate context. They are based on technologies such as HTTP protocol, URI and RDF. The next chapter will discuss the four distinctive solutions that implement the concepts exposed in different applications.

3. SHOWING PLATFORM FOR THE PRESENTATION OF OPEN DATA

Today, the market can find dozens of platforms to view and search abstracts of open data. Commercial solutions are mostly based on known platform for documentation management (ECM / DMS) and represent their extension. Our approach to the representation of a set of open data should be used platforms are open source.

In this paper, we will make an analysis of four very popular platform for presenting open data that are open source. Will present their comparative characteristics according to the criteria and principles we have presented in the previous section. Finally, we propose a solution that would be suitable for use in our practice.

The choice of solutions that will be analyzed based on the following criteria:

- platform is open source
- have filed applications that include the following data owner: used in ministries, state agencies and local governments (refers to platforms that are applied in the public sector) used by universities and institutes for the presentation of scientific papers, the results of research and teaching support
- number of installations is greater than 20
- provide further linking and aggregation of data - meet the class of minimum 3 stars

We analyzed the following platforms for presenting open data sets:

- *CKAN* platform for publishing, sharing and retrieval of data, which are used as the basis for detailed information in the catalogs of open data.
- *Ambra* is a platform for the presentation of data from scientific studies and articles
- *Google Fusion Tables*, a solution to easily publish and integration of open data
- *OGDI* platform to display data based on open. NET technologies

	Name	URL *()*
1	CKAN	*http://* ckan.org
2	Ambergris	*http://* ambraproject.org
3	Ogdi	*http://* datapublic.org
4	Google Fusion Tables	*http://* tables.googlelabs.com
	Legend:	
	K1	Application: The G-public sector, science-, ST-statistical indicators of
	K2	Number of installations
	K3	Support any ways to access data from a set of open data: F - the search for facts I - from data to information (reports, visualization) C - from the data interface (combining data sets) E - from data to information (export to various schemes) S - the data to the service (integration with existing and create new ones)
	K4	Model 5 star: number of stars - the degree of associated data

Table 3 Comparative platform to display a set of open data

CKAN. Nowadays the most widely accepted solution for the publishing of open data. The solution is based on the Python language and program JavaScript to display on the user side. Use the Pylons web framework and SQL Alchemy as a ORM, and data storage Postgres SQL database. The architecture allows easy and rapid development of new extensions that bring new functionalities and features for working with data. The solution is very well covered in the documentation. Supports all modern concepts that characterize open data. It is used by over 100 institutions including ministries and state government agencies G8 countries such as the USA, Russia, UK, Germany, France, Japan, Canada. A favorite

and local governments, cities, which include Berlin, Houston, Washington.

Ambergris. Ambra is an innovative Open Source platform for publishing the results of scientific research. It provides opportunities for publication of notice and hearing, which allows a "living" document around which the possible scientific discussion and consideration of new scientific discoveries. The platform is under active development by PLoS (Public Library of Science) and licensed under the Apache License, Version 2.0. The platform is developed in Java, using Spring, Struts, and in the middle layer Hibernate. As the application server uses Apache Tomcat platform and MySQL database. Key users Ambra platform for the presentation of open data are scientific institutions gathered around PLOS ONE project (http://www.plosone.org/). The project brings together over 50 institution and over 100 collections of open data. Platform meets the criteria for linking open data class 4 stars.

Google Fusion Tables. This solution is based on cloud services. It is well connected with all the Google tools. Data can easily publish and intuitive way to manage the published data. The solution has a strong visualization (maps and charts), the ability to work with large data sets and integration with other Google services. Ease promoted by Google impact that this decision is not supported by any strict concepts of open data, and that certainly meets the 3 star class, and that the higher the level of integration of data coming applicative upgrades to the potential users themselves do. Number of active users of this solution is great. Ability to readily publication and presentation of data makes this solution a good extension to connect with the existing portal-user solutions to public sector and science.

Open Government Data Initiative (ogdi) is an open source cloud solution that promotes the use of open data and provides assistance to a faster and easier way to publish data sets [8]. Data can be displayed in several different formats such as tables, maps, charts. ODGI is based on Microsoft. NET platform, written in C # using the Windows Azure platform. Many datasets, including geospatial data, which returned KML format, make ogdi compatible with popular desktop and web-based technologies such property, such as Microsoft Bing Maps, Google Maps, Yahoo Maps and Google Earth. Ogdi solution is easily connected to a Microsoft portal-products, so that a fast open public sector data for users who rely on Microsoft technologies. The number of users of this solution is dominant in the United States and Canada, primarily in open data local administrations are the traditional users of Microsoft Share Point platform.

4. RECOMMENDATIONS FOR THE SELECTION OF THE PLATFORM

Practices

In the middle of 2013th all the states of the G8 and the G20, and EU member states have opened their data. In numbers it looks like this: over 250 open data portal, which are the responsibility of the state (includes portals of ministries and agencies), region, local governments. Of the total number of portal shall collect 39 federal states in the U.S., 41 cities and areas in the United States, 43 are under the control of the states in the world, and 160 cities and regions in the world [9]. Data from the June 2013th mostly to open more than one million different datasets.

The analysis of data usage openly *(source: data.gov),* reveals a set of dominant areas (data sets) from which the data is still the most in demand (Figure 3):

- geospatial data
- datasets on transport,
- agriculture,
- energy and energy services,
- various statistical data,
- data relevant to scientific and technological development.

Figure 3 shows the number of downloads data from the portal (the average number of monthly, data.gov portal) and the use of data sets in practice.

Practical experience shows that the portals solutions to openly display data based on CKAN platform when a set of data relating to the data and state government agencies, as well as amber and Google Fusion Tables, which are dominant display data resulting from projects and scientific research.

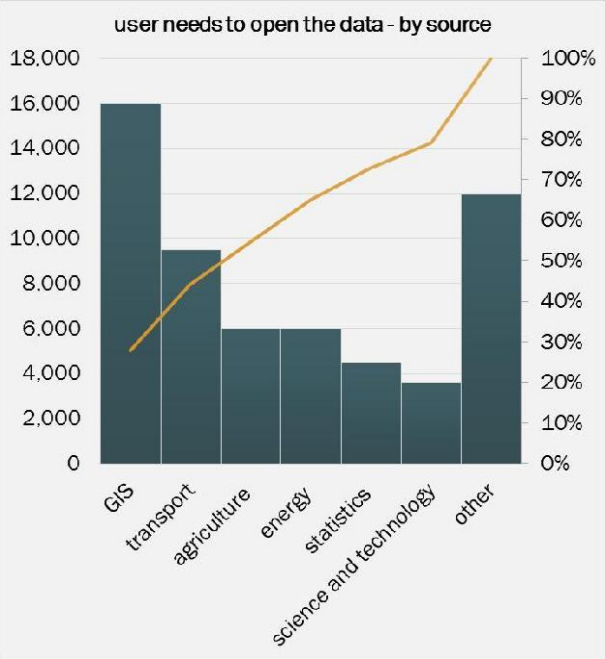

Figure 3 Downloading data from a set of open data by the user (data.gov)

Recommendation platform suitable for the public sector

Our recommendation is to CKAN taken into immediate consideration when choosing a platform for open display of data in the public sector. This platform is still the most mature and complete, with a variety of installation and implementation, and there are multi-year plans for the development of new functionality. Subscribe to a large community of users and a large knowledge base with the experiences of practical use that can measure dozens of man-years. The platform is open for extension and integration with existing back-end systems of institutions.

Implementation risks CKAN platform in practice is reduced to solving the technical problems of integration with the data generated in the existing information systems of institutions.

Recommendation Platform for Science and Education
As a recommended platform for the presentation of open data in scientific research institutions believe that it is possible to use three of these platforms, depending on the type of institution and the scope of the material to be presented.
If we look at the level of **scientific laboratory** / faculty **chairs** - *Google Fusion Tables* are the best choice. The platform is easy to integrate in existing portals laboratories, departments, or projects, the data can be easily published and she has a rich set of tools for visualization [10]. This solution is particularly useful when we want students and key persons to present the most significant achievements of research.
When presenting data from **large scientific research projects** *AMBRA* is right. Datasets are presented in AMBRA solutions are separated by project, includes versioning results. Composing the social integration of several related projects participants and beneficiaries of the research results (companies) make it possible that the discussion comes to quality and detail to explain the conclusions obtained. Sponsors research tracked discussions and actions impact on investment. Flow analysis discussion better define the priorities of the next cycle of research projects [11].
If the process of opening up data associated with scientific research to look at the level of institutions such as the Ministry of Science, Scientific Research Institute, the University or the University as a whole, then the CKAN solution of choice. We suggest this platform because of the types of documents that should be included in the process of opening, and that the addition of the results of research and status documents institution, decisions and so on.

5. CONCLUSION

Signed agreements with organizations in the Open Government Partnership mart 2012. The Government of Serbia has started the process of developing strategies related to open data.
This paper indicates the steps to be followed in the process, which includes identification of data sets that need to be opened, the legalization of the use of open data formats and services that open data made available to users, and the choice of applicative platforms that can be used in the presentation of open data.
The data themselves do not have any particular value. The publication of data is useless if the data is not used to solve real social, scientific and business problems. Empowering the community, citizens and businesses, the process of opening the data means focusing on the use of data and problems to be solved that way.
Prevent anyone datasets should be done according to the priorities and dynamics that dictate the use of this information in practice. This would allow the following benefits which would be the direct beneficiaries of the citizens and the economy:

- Increased confidence of citizens and businesses to government agencies and other institutions of the state, as well as its service
- Further development of entrepreneurship, especially the service sector that could efficiently and effectively use open data to improve existing and develop new services
- Increase agility government organizations and effective connectivity with the market
- Analysis and available results of scientific research that is funded from the state budget, which would continue to provide businesses and entrepreneurs opportunities to develop new products and services.

The process of opening the data may not represent the activity focused solely on technology, and not focused on the data itself.
The process of opening the data to focus on the goal of improving the overall quality of life: reducing pollution and resource consumption, improved health care, better connectivity of people, An increase in the productivity and reduce operating costs - a total increase of living standards of people.

REFERENCES

[1] Miller, P., Styles, R., & Heath, T. (2008, April). Open Data Commons, a License for Open Data. In LDOW.

[2] Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project, Government Information Quarterly, Volume 30, Issue 4, pp. 508-513

[3] Open Data Handbook Documentation. (2012, Novembar). Open Knowledge Foundation. Retrieved from http://opendatahandbook.org /pdf/OpenDataHandbook.pdf

[4] Zuiderwijk, A., & Janssen, M. (2013). Open data policies, their implementation and impact: A framework for comparison. Government Information Quarterly.

[5] Murray-Rust, P. (2008). Open data in science. Serials Review, 34(1), 52-64.

[6] Berners-Lee, T. (2006). Linked Data. International Journal on Semantic Web and Information Systems (IJSWIS), 4(2), 1.

[7] Berners-Lee, T., et. al. (2006), Tabulator: Exploring and Analyzing Linked Data on the Semantic Web. Procedings of the 3rd International Semantic Web User Interaction Workshop (SWUI06).

[8] Miladinović, M. (2013). Otvoreni podaci u javnom sektoru. Specijalistički rad. Visoka tehnička škola Kragujevac.

[9] Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far.International Journal on Semantic Web and Information Systems (IJSWIS), 5(3), 1-22.

[10] Stefanovic, M., Matijevic, M., Cvijetkovic, V., & Simic, V. (2010). Web based laboratory for engineering education. Computer Applications in Engineering Education, 18(3), 526-536, doi:10.1002/cae.20222

[11] Hester, J. R. (2013). Closing the data gap: Creating an open data environment.Radiation Physics and Chemistry.