What is Open Data
and How to Benefit from It

Sebastian Grabowski

Orange Polska S.A
Centrum Badawczo-Rozwojowe /Research and Development Center

Abstract

Purpose/Thesis: The aim of this paper is to introduce the concept that Open Data and Open APIs provided by Communication Service Providers integrated in one end-user-oriented application may considerably improve the process of communication between people and institutions.

Approach/Methods: Open Data is one of the key elements of the broad Internet ecosystem; other elements, such as open interfaces, open source, API, etc., are the assets that make the Internet environment robust, scalable, and extendable. The paper, based on the case study analysis, presents selected applications integrating the communication enablers in the form of Open APIs and Open Data sources.

Results and conclusion: The combination of Open Data and functions provided by telecommunications operators in the form of Open APIs significantly improves and facilitates the processes of communication between people and institutions.

Originality/Value: The author proposes to integrate Open Data with real time communication functions provided by Communication Service Providers in the form of Open APIs. Open Data and Open APIs are effective tools to create user-made environments that are convergent and coherent from the application source code point of view.

Keywords

Open Data. Open API. Service interfaces. Telecommunications operators.

Received: 21.03.2014.   Revised: 8.05.2014.   Accepted: 16.05.2014.

1. Introduction

Open Data is one of the symbols of our times. Its nature and application is best defined by DBpedia (Auer, et. al. 2007, 722):

> Open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control.

Another definition was created by the Open Knowledge Foundation in 2005 (Defining Open Data, 2013):

> Open data is data that can be freely used, shared and built-on by anyone, anywhere, for any purpose.

It seems conclusive that Open Data should be easily accessible to all citizens without any legal or technical limitations. What purpose does Open Data serve? First and foremost, it facilitates the transparency of public life. When combined with information from web

portals, social networking sites and facilities provided by Open API, Open Data enables internet users to create new innovative portals, applications and systems. Open Knowledge Foundation portal defines three key features of openness (Defining Open Data, 2013):

> Availability and access – the data must be available for everyone in convenient and modifiable form.
> Reuse and redistribution – terms and condition should permit reuse and redistribution of data in machine-readable form.
> Universal participation – everyone must be able to use data sets.

Open Data significantly influences three areas: open governance, smart cities and innovation.

In open governance Open Data showing politicians' activity gives citizens the opportunity to evaluate their representatives in the process of governance. The information about the results of parliament voting , absences and activity in the mass media can be very useful for the assessment of politicians.

Smart cities is another large area of potential Open Data application. The smart city concept is focused on providing models and specific solutions based on innovative new technologies, new organizational models and management as well as new legal regulations for the development of urban infrastructure, efficient resource usage, support of social interactions and construction of social capital (Muraszkiewicz, 2014). Open Data is extremely important here as it may be used to collect and expose information about: the activity of citizens on the city portals, housing prices, traffic information, safety, number of students in schools, energy and water consumption, etc.

In the case of innovation Open Data refers to the concept of Open Innovation by Professor Henry Chesbrough (Chesbrough, 2006). Open Data exposed with API allows the users to create innovation (e.g. new innovative applications and services) outside the place (organization) where the data is stored.

In this paper the author focuses on the examples of applications based on the concept of Open Data and Open API exposed by telecommunications (telecom) service providers – in particular on the features provided by Orange. It should be noted that the legal aspects of the selected applications discussed here have been limited to the case of the telecommunications market in Poland.

2. How to make Open Data accessible?

According to the definition quoted in the previous section, Open Data is assumed to be readily available for reuse and reprocessing. In order to meet these conditions, data providers are required to expose data in such a way as to make it:
  – easily accessible for potential users,
  – available for a number of users simultaneously,
  – regularly updated,
  – secure in terms of access, both for the supplier and the receiver.

It seems absolutely necessary for the entity providing the data to build a platform that will be able to expose data in a safe and productive way. The first issue to be taken into consideration is the architecture of the platform since the system may be built with either

centralized or distributed architecture. The centralized option ensures the security of the Open Data. In that case all hardware and software elements of the platform may be installed in the data center of the institution exposing the data. On the other hand, it is a considerable challenge to make such an environment productive and fully available. The performance may be improved mainly through the enhancement of the resources (that is via additional servers, databases, network devices), which is an expensive solution. Another approach is a distributed system. In this case the elements of the system may be situated in different places and with various entities being located in so-called *cloud*. As a result, Open Data is available in different places in the networked environment, which on the one hand solves the problem of performance, but on the other hand makes the users face the challenge of keeping stored data up-to-date in the distributed environment. This requires the development of tools for synchronizing data with the primary source.

Another matter to be considered is the compliance with the standards and IT protocols used for the data exposure. In order to make data reusable and easy to use in web portals, applications and systems, the programming interfaces known as API (Application Programming Interface) should act as data carriers. The data repository using the API can provide the end user with the data represented via the most popular Internet protocols, such as HTTP (Hypertext Transfer Protocol), FTP (File Transfer Protocol) or LDAP (Lightweight Directory Access Protocol). These protocols usually send text data. However, the user may also take advantage of less popular protocols (e.g. binary protocols) which are more difficult to implement but offer better security and compression of transmitted information.

The contemporary trends prove that in the majority of cases the platforms exposing both Open API and Open Data use Web Services technologies based on HTTP protocol. However, another issue arises at this point, namely Web Services (WS) exist in two different models. One of them is strictly related to Service Oriented Architecture (SOA) in the case of which web services take advantage of SOAP protocol and service semantics written in WSDL specification (Newcomer, 2003). Slightly simpler attitude is proposed by the RESTful architectural style (Richardson et. al. 2013) which is based on the Resource Oriented Architecture (ROA), where the access to the service (e.g. to the Open Data) is gained through URL (Uniform Resource Locator) resources and methods in HTTP protocol (such as GET, POST, PUT, DELETE).

Another topic demanding attention appears here, i.e. the format of the data itself as exposed by Web Services. Simple data can be displayed in a text format, e.g. Comma Separated Values (CSV), or in the key-value format, but also with the help of XML (Newcomer, 2003) and JSON[1] (Rockford, 2009) notations.

When dealing with complex data types containing, for instance, graphics, the way of exposing such content by Web Services must allow for the simple download of such objects. The instances of such data types are the Geographic Information System (GIS) records representing maps which are provided via e.g. Web Map Service protocol (WMS).

The application of open protocols and standards presented in this section facilitates both the exposure of Open Data and the creation of applications using it. The accessibility of SDK sets (Software Development Kit) accelerates and simplifies the job of the programmers even further. SDK modules, available for particular programming environments (Java,

---

[1] http://www.json.org

MS Visual Studio, PHP, etc.), provide programmers with objects that may be processed by classes and application methods. Using the objects of Enterprise JavaBeans specification of Java language is very good example of this type of high-level programming.

3. The sources of obtaining Open Data and potential problems

The most crucial sources of Open Data are governmental and educational institutions. Governmental institutions „effortlessly" acquire some collections of citizen data, industrial data, statistical data, etc. Some of this data is sensitive because of the Polish law on the Protection of Personal Data and it cannot be made public. However a considerable number of records (like statistical data) may be published within the Open Data model. The examples of data repositories coming from Polish governmental sources are: Otwarty Budżet (Open Budget) and Sejmometr (Parliament-Meter) sites.

Another substantial collection of Open Data is the data related to the educational institutions. In 2004 the education ministers of all member nations of the Organization for Economic Cooperation and Development (OECD) signed a document in which they declared that all data obtained as a result of the allocation of public funds must be publicly accessible. The significant number of projects and communities on the Web is dedicated to the open publication of data. It is worth mentioning here the chemistry-oriented Blue Obelisk group[2] and Linked Science project[3].

It should also be pointed out that there are two other potential sources of data (perhaps still not treated as Open Data), that is, social networking and telecom operators. Social networking sites such as Facebook, Twitter or Google practically base their entire business model on acquiring data from the Internet users and making them available to third parties (mainly in order to present contextual advertisements to the users).

Another issue is the acquisition of data from the public, e.g. applying the methods of crowdsourcing. From the end-user point of view it is essential to provide a convenient interface for entering data or making it available. These interfaces can be offered in the form of:
  – websites,
  – dedicated desktop applications,
  – mobile applications.
Two first mentioned above include, for instance, portals enabling citizens to report problems with infrastructure, such as FixMyStreet[4], FixMyTransport[5], or the Polish initiative Naprawmy To![6] (Let'sFixIt!). The latter also offers a mobile application for Android (operating system). Another interesting example of a mobile application used to capture and share information within Open Data model is the application called Yanosik[7]. This system warns users against the police patrols, speed cameras and traffic incidents and all information can be entered into the system by the application users.

---

[2] http://www.blueobelisk.org/
[3] http://linkedscience.org/
[4] http://www.fixmystreet.com
[5] http://www.fixmytransport.com
[6] http://naprawmyto.pl/home
[7] http://yanosik.pl/

4. Open API – the key to success

Programming interfaces have been well known in the computer science for many years. The adjective "open" has already been used in the names of numerous computer interfaces – just to mention Microsoft's ODBC (Open Database Connectivity) or OpenGL standards of graphics programming. The term "Open API" appeared for the first time in the Internet world of social networking and Web 2.0. The creators of the term meant to emphasize the ease of combining different platforms and systems based on Web Services. It is important to mention that Open APIs are published on the Internet and their specifications are available to all potential users. Why Open API is so crucial? First of all, it enables programmers to create interesting end-user applications of commercial value. The "interesting" application should have the following characteristics:
- combine Open APIs from various sources (mashup),
- use or gain Open Data,
- have an attractive graphic representation,
- have rich functionality,
- have low hardware requirements,
- in the case of mobile applications provide low power consumption in relation to the battery drain,
- have a "reasonable" price.

Combining all, often even contradictory, features listed above, in one application is not easy. The large number of free applications available on the web have not proved to be commercially successful and few of them deserve to be called "killer applications".

Telecom operators have also managed to recognize the potential of Open APIs and made a number of features available for the programmers on the Internet. Some of those functions are: receiving and sending SMS, MMS or USSD (called Unstructured Supplementary Service Data), locating mobile terminal (mobile phone), checking its status (free, busy) or adding amounts for goods and services purchased on the Web to the subscriber's account (payment API).

5. Examples of applications using Open APIs and Open Data

Orange Labs created a number of applications within the Open Middleware 2.0 Community project that integrate functions of a telecom operator with an Open Data facility (Podziewski et. al. 2012; Bogusz et. al. 2012; Litwiniuk et. al. 2012), and the concept of using an Open API in a variety of systems and applications (Kalitska, et. al. 2012; Legierski & Korbiel, 2011). This section depicts several applications that use Open APIs and Open Data.

Emergency Button (Podziewski et. al. 2012) – the idea of the Emergency Button application is based on the interaction between the senior and the caregiver. Seniors are people with health problems requiring special care, help or assistance. If they use mobile phones, they are given an opportunity to start the Emergency Button service (by pressing the key with programmed USSD code or automatically – whenever the phone detects the user's fall). When the action is initiated, the caregiver receives a message (an SMS) containing the location of the senior including his/her geographical coordinates and approximate

address. The application uses the following API: Receive USSD, SMS Send and Terminal Location, as well as the Google Maps API.

BusStop (Litwiniuk et. al. 2012) – this is an application focused on public transport. The system is to provide the user with the information on the timetables of selected bus lines, the routes of individual lines and lines running in the user's vicinity. This service is based on the functionalities offered by mobile networks. BusStop application uses USSD channel through which it is possible to provide the service. In the next step, the system locates the user and his nearest bus stop or selects the timetable required by the user. In the final stage, the information requested by the user is sent to him via SMS. This API application employs: Receive USSD, Send SMS, Terminal Location and the data from ZTM Warszawa (Warsaw Public Transport Authority).

Telco 2.0 for UC (Bogusz et. al. 2012) – Unified Communications is a concept which has recently become popular on the market of business communication. It combines all possible communication channels in one application and for one phone number. The UC communication can be expanded with such channels as, for instance, SMS, MMS and USSD. Telco 2.0 for UC enables the use of the operator function in the system from Siemens – Openscape UC. The typical user has his office VoIP phone connected to the OpenScape Voice software telecom switch using the SIP protocol. On the other hand, such a subscriber has access (using a computer and web browser) to the OpenScape Web Client application, which enables the use of advanced Unified Communications functions such as conferencing, Instant Messaging conversations, phone management (CTI – Computer Telephony Integration), etc. The Architecture of the Siemens OpenScape environment has been complemented with the Telco 2.0 for UC server, the task of which is, on the one hand, to establish and maintain a communication session (using Open API) with the service provider's Service Delivery Platform (SDP), and, on the other hand, to expose services to the subscriber of the Unified Communications system. The Telco 2.0 for UC allows the users to send SMS and USSD messages, as well as locate the mobile terminal. The application uses Send SMS, Send USSD and Terminal Location APIs. Other Open APIs used in the system are Google Maps and OpenScape UC SOA API.

Social Game – This is a mobile social networking city game. The main objective of this scheme was to build a platform which enables users to create their own scenarios for urban games. The scene of the game can be any place specified as its territory by the designer of the game. The player can be anyone with a mobile phone offering access to the Internet. The audience passively using their laptops or desktops transfer hints to the players.

The idea of an urban social game is based on the use of information provided by the operator's platform which exposes Open APIs. As a result, the city game has been complemented with information provided directly by the mobile network. This additional telecommunications aspect will allow for the use of a mobile phone as a terminal to control the game in which the users participate. Through USSD messages the users can be informed about events taking place in the game which are significant from their point of view, and they may inform the platform about their activities. Additionally, the system specifies the player location via the platform in order to handle the scenario. The application uses Send SMS Terminal Status, Terminal Location and Facebook API.

6. Telecom operator as a potential source of Open Data

Telecom operators store extensive amount of data both about their customers and telecommunications infrastructure. Some of it can be published and made available as Open Data. This type of data includes, for example, information about the architecture of the mobile network. Data such as the location of BTS (Base Transceiver Station) and information about the mobile cell (Cell ID) are the subject of two crowdsourcing projects, that is BTS Search[8] and Open Cell ID[9]. The information concerning the infrastructure components of those mobile operators is collected by the Internet users and made available as Open Data. However, the telecom operator is able to expose the same set of data on the Internet in the form of Open API. It should be emphasized that this data will obviously be more current than the information gathered by the Web community.

The operator can even go a step further and offer more sophisticated and detailed data (this time, for a fee) including e.g.: the location of the BTS, the Cell ID values, Base Station Identity Code Neighbor Cell and information about the ID of BCCH (Broadcast Control Channel ID). This data set can be useful for external entities to build algorithms determining more suitable location of mobile stations and evaluation of the quality of the network operator.

Another set of data that can be offered in the form of Open Data may represent aggregated information about telecommunications traffic between cities. When selling data, the provider can offer more detailed set of information on, for instance, the telecommunications traffic observed from the BTSs.

To sum up, it seems conclusive that the telecom operators should not oversee the benefits of data sharing both in the Open Data and paid data models and develop business models in which they can offer these types of data to their customers.

7. Conclusions

The use of Open API programming interfaces is the prerequisite for the effective use and re-use of Open Data. It is strictly related to the fact that the productive use of data basically means its employment in web applications and web portals. Only the exposure of data in the form of API by the source system can make it up-to-date and valid for the user. At the same time, it can secure the data and protect the integrity of data transmission.

When talking about the exposure of Open Data and Open API in recent years, in particular in the Internet environment, it is possible to observe trends in the use of Web services based on RESTful protocol and data representation in the form of XML and JSON standard. Standards related to the use of SOA and SOAP protocol, also being upward trends, have been popular mostly with private networks of companies and government organizations.

The examples of applications created in Orange Labs clearly show that the combination (mashup) of Open API and Open Data is crucial for the functionality of innovative applications, and the Open Data itself made available by means of Open APIs is a key element of the network ecosystem.

---

[8] http://btsearch.pl
[9] http://www.opencellid.org/

References

Auer, S. R.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. (2007). DBpedia: A Nucleus for a Web of Open Data. In: Aberer, K., et al. (eds.) The Semantic Web. 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11–15, 2007. Proceedings. Berlin: Springer, 4825 722–735.

Bogusz, D.; Legierski, J.; Podziewski, A.; Litwiniuk, K. (2012). Telco 2.0 for UC – An example of integration telecommunications service provider's SDP with enterprise UC system. In: M. Ganzha, L. Macaiszek, M. Paprzycki (eds.) *Computer Science and Information Systems (FedCSIS) Proceedings*. Warszawa: Polskie Towarzystwo Informatyczne, IEEE Computer Society Press, 603–606.

Chesbrough, H. (2006). *Open innovation*. Boston: Harvard Business School Press.

Defining Open Data. (2013) [online]. Open Knowledge Foundation Blog, [24.04.2014], https://okfn.org/opendata/

Kalitska, S.; Kukielka, P.; Jonczyk, M.; Legierski, J.; Szczekocka, E. (2012). Forecasting of threatening situations in Smart Space. In: M. Ganzha, L. Macaiszek, M. Paprzycki (eds.) *Computer Science and Information Systems (FedCSIS) Proceedings*. Warszawa: Polskie Towarzystwo Informatyczne, IEEE Computer Society Press, 641–647.

Legierski, J.; Korbel, P. (2011). Telco 2.0 – przykłady praktycznego wykorzystania interfejsów telekomunikacyjnych platform usługowych [online], [30–05.2014], http://www.tu.rd.tp.pl/portal/article/Telco_2.0_KSTiT_2011_Przyklady_implementacji.pdf

Litwiniuk, K.; Czarnecki, T.; Grabowski, S.; Legierski, J. (2012). BusStop – Telco 2.0 application supporting public transport in agglomerations. In: M. Ganzha, L. Macaiszek, M. Paprzycki (eds.) *Computer Science and Information Systems (FedCSIS) Proceedings*. Warszawa: Polskie Towarzystwo Informatyczne, IEEE Computer Society Press, 649–653.

Newcomer, E. (2002). *Understanding Web Services: XML, WSDL, SOAP, and UDDI*. Boston: Addison-Wesley.

Podziewski, A.; Litwiniuk, K.; Legierski, J. (2012). Emergency button – A Telco 2.0 application in the e-health environment. In: M. Ganzha, L. Macaiszek, M. Paprzycki (eds.) *Computer Science and Information Systems (FedCSIS) Proceedings*. Warszawa: Polskie Towarzystwo Informatyczne, IEEE Computer Society Press, 663–677.

Richardson L.; Ruby S. (2007). *RESTful Web Services*. Farnham: O'Reilly.

Open Data – czym są i jak z nich czerpać korzyści

Abstrakt

Cel/teza: Celem pracy jest przedstawienie koncepcji mówiącej, że Otwarte Dane i funkcje dostarczone przez operatorów telekomunikacyjnych w formie Open API zintegrowane w aplikacji użytkownika końcowego mogą w sposób istotny poprawić proces komunikacji pomiędzy ludźmi i instytucjami.
Koncepcja/metody badań: Otwarte Dane są jednym z kluczowych elementów ekosystemu Internetu. Inne elementy, takie jak: otwarte interfejsy programistyczne, open source, API sprawiają, że środowisko Internetu jest niezawodne, skalowalne i rozszerzalne. W publikacji przedstawiono przykłady zastosowania aplikacji integrujące funkcje komunikacyjne w formie Open API i Otwarte Dane. Praca jest oparta na metodzie studium przypadku (case study).
Wyniki i wnioski: Głównym przesłaniem jest konkluzja, iż Open Data w połączeniu z funkcjami dostarczanymi przez operatorów telekomunikacyjnych w formie Open API znacznie poprawiają i ułatwiają procesy komunikacji między ludźmi i instytucjami.

Oryginalność/wartość poznawcza: W niniejszej pracy zaproponowano integrację otwartych danych z funkcjami komunikacji w czasie rzeczywistym świadczonymi przez operatorów telekomunikacyjnych w formie Otwartych API. Przykłady aplikacji pokazują, iż Open Data i Open API to proste, ale skuteczne narzędzia do tworzenia przez użytkownika środowisk aplikacyjnych, które zapewniają konwergentność i spójność kodu.

Słowa kluczowe

Otwarte Dane. Otwarte API. Interfejsy usług. Operatorzy telekomunikacyjni.

SEBASTIAN GRABOWSKI – Graduated from the Faculty of Electronics and Telecommunications at Koszalin University of Technology, the Faculty of Management at the University of Warsaw and the Warsaw School of Social Psychology. His research focuses on Open Data, Open API, interactions between users of telecommunications networks and the Internet as well as issues related to Open Government. For the past 12 years he has been involved in the telecommunications industry, currently he works in Orange Poland. He is the Director of Orange Poland Research and Development Center. He is a co-founder of the Open Middleware 2.0 community – a program bringing together different communities of developers, researchers, technology suppliers and IT companies near the world of Telecommunications and IT.
Contact to the Author: Sebastian.Grabowski@orange.com
Orange Polska
Badania i Rozwój Sieci i Platform Usługowych
Centrum Badawczo-Rozwojowe
Obrzeżna 7
02-691 Warszawa