

PUBLICATION AND REUSE OF OPEN GOVERNMENT DATA

A Practical Approach

T. Cerdeña Hernández, F. Fumero Batista, L. M. Moreno de Antonio

D. Pérez Barbudo and J. L. Roda García¹

Escuela Técnica Superior de Ingeniería Informática, Universidad de La Laguna, Tenerife, Spain

Keywords: Open government data, Gov 2.0. Transparency, Linked data, Innovative applications.

Abstract: Web 2.0 has changed the way that information is presented to people. Public administrations have an important role to play in this new era. Global institutions, central, regional and local governments gather and produce a wide variety of information that is potentially reusable by citizens and the digital content industry. Gov 2.0 follows the approach of giving access to open public data to citizens. Open Government Data (OGD) establishes the principles for providing public data to the public. The development of innovative applications by companies or individuals from these public records, will meet the demand of information from citizens, as well as developing the basic principles of transparency, publication and reutilization. We present a local administration experience showing the steps to implement the OGD strategy. Benefits and responsibilities of those involved in the full process are presented. We have developed a real case of publicly available data from the population census of the municipality.

1 MOTIVATION

Public administrations gather and produce a wide variety of information that is potentially reusable by citizens and the digital content industry. Public Sector Information (PSI) is a valuable resource for the society (European Directive 2003/98/EC, 2003).

Web 2.0 (O'Reilly, 2005) has meant a significant change in terms of individual and joint contributions of the users: blogs, wikis, social networks, etc. We can say that Government 2.0 or Gov 2.0 is the use of technologies used in Web 2.0 to solve problems collaboratively among citizens and institutions (Lathrop and Ruma, 2010).

Every day, more websites offer OGD. Among the most representative we can mention countries like UK (data.gov.uk), USA (data.gov), or institutions like the United Nations (data.un.org) or the World Bank (data.worldbank.org). On the website of the U.S. government (<http://www.data.gov/community>) and the CTIC of Asturias, Spain (<http://datos.fundacionctic.org/sandbox/catalog/faceted/>) we can find information from the various institutions that offer catalogues of public datasets. It

is important to consider that the institutions must provide reliable, actual and robust data, so companies and individuals can develop innovative applications for the web or the mobiles.

In section 2, OGD is presented. The developing experiences, the architectural structure and technologies of the publication and reusable systems are presented in Section 3. We finish with the conclusions and the references.

2 OPEN GOVERNMENT DATA (OGD)

In December 2007, different people and groups interested in OGD met and postulated eight basic principles for OGD (Open Government, 2007). The data must be complete, primary, timely, accessible, machine process, non-discriminatory, non-proprietary and license-free.

OGD (Berners-Lee, 2010) is a strategy that enables publicly available government information to be easily accessible in order to develop new innovative applications for the interest of citizens.

¹Authors are listed in alphabetical order. This work has been partially supported by the FEDER and the Agencia Canaria de Investigación, Innovación y Sociedad de la Información (ACIISI), Gobierno de Canarias (ProID20100252).

This strategy must offer information in open and free formats and with the security and trust needed.

The benefits and responsibilities of OGD strategy can be seen from three role viewpoints: citizens, administrations and infomediaries. Citizens can obtain free and open access to information collected and generated by the public sector, no proprietary formats. Citizens can see institutions more transparent, and therefore some control of them. For the application developers (infomediaries) they can develop new products and services, promote new innovative solutions, and improve their business. Public administrations can obtain citizens feedback on the information they published, and reduce time, costs and efforts to provide information to the citizens.

Not only benefits but also responsibilities are achieved for all the participants. Citizens primarily must collaborate and offer feedback to administrations. Infomediaries must use new technologies that improve the way administration workers and citizens' access to data. The public administration must provide transparency, must provide a broad spectrum of data and must offer a stable and robust platform for present and future.

Currently, one of the most important aspects in OGD is the format of public data. There are two main trends at least. On one side, public administrations web sites offer great unlinked data files in proprietary formats or open (Excel or CSV). The principal problem is that integration with other data and systems is no easy. On the other hand, some institutions seek to use data linked through to linked data-based formats such as RDF (RDF, 2010) or ontologies (OWL, 2010).

There are several works that provide a set of steps to develop OGD strategy (O'Reilly, 2010):

1. Develop the policy directive ((www.sfmayor.org/wp-content/uploads/2009/10/ED-09-06-Open-Data.pdf)).
2. Create a simple infrastructure, reliable and offer publicly accessible public data.
3. Try to use open standards that allow interoperability with other systems.
4. Create web sites that present the catalogue of data and develop some applications.
5. Share API developed with citizens to gain access to data without the direct intervention of the institution.
6. Share work developed with institutions in other countries, regions and Municipalities.
7. Create a list of applications that can be reused by employees of the institution.

8. Create an app-store to accommodate all public or private applications.

9. Encourage citizens and businesses to develop applications.

10. Create communication channels that allow citizens to make proposals for new applications from the data shown publicly.

11. Instruct both employees of the institution and citizens.

There are different technical specifications related to the Semantic Web that are used in OGD. The Resource Description Framework specification, RDF (www.w3.org/RDF/) is a standard for exchanging data on the Web. RDF allows the unique resource description in the web space and establishes relationships with other objects. You can execute queries on these data through the SPARQL language (www.w3.org/TR/rdf-sparql-query/). We can find different tools to convert from various formats to RDF in (esw.w3.org/ConverterToRdf).

3 A CASE STUDY

3.1 Municipality Population Census

San Cristóbal de La Laguna is the third most populated municipality in Canary Islands, Spain (<http://www.aytolalaguna.com>). The Gerencia de Urbanismo is an autonomous institution that depends from the town hall and takes care of many different services related to urbanistic issues. The Population Census is one of the most important sources of citizen's information.

3.2 Requirements

The release of public information of interest for the citizens is becoming a major concern for the IT responsible at La Laguna local government. The purpose of this experience is to prove the feasibility of OGD implantation in this administration. Later on, and based on the experience acquired while developing this example, we would develop a well-defined protocol to follow in order to extend the release to other datasets. Besides this, there were other ideas to take into account:

- Data to work with: use existing real data to face real issues, and work with a small but representative dataset.
- Integration with existing systems: keep current data stores isolated from the publication system for safety and integrity, allow to publish from

heterogeneous data sources, and minimize the impact on current data stores

- Reutilization: publish data in a format easy to use for reutilization, and develop a sample application of reutilization for the published data.

It is important to remark that the data provided was clean of privacy personal information even before we started working on it. The data consists on a list of properties within the municipality boundaries. Each property in the list has an address (in the street system of the municipality), a location (latitude, longitude) and the quantity of inhabitants (only those registered in the census).

3.3 Proposed Architecture

Our proposal was to develop a system that fulfilled the requirements of the administrators while following the steps suggested in section 2.

Administrations store a huge amount of information in many heterogeneous formats, such as relational and non-relational databases from different vendors, spreadsheets, documents, texts, etc. Linked Data is the chosen solution in this example to represent this information in a common format.

The designed architecture is showed in Figure 1 and Figure 2. It is composed of two main systems; each of one is explained in the following sections.

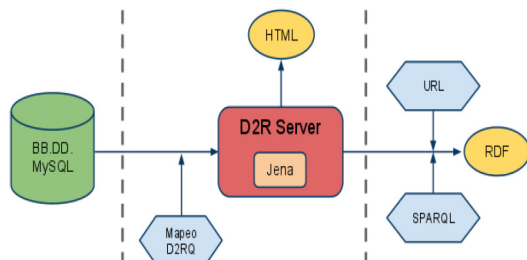


Figure 1: The Publication System.

3.3.1 Publication System

The main component of the publication system developed for this example is an instance of a D2R server (www4.wiwiss.fu-berlin.de/bizer/d2r-server/). The D2R server allows publishing data from relational databases as RDF. In this example the source data from the census is stored in two tables in a MySQL server instance, one for the properties and one for the streets. The main task here is to define a vocabulary to properly represent this data in RDF and then create the corresponding mapping in the D2R server. This way the data is immediately exposed in an open reusable format.

D2R server also allows making SPARQL queries to the database which improves the reutilization possibilities.

3.3.2 Reutilization System

The reutilization system is built as a separate instance to expose an example of reutilization of the information published in RDF. This example is a web application which displays census data geolocalized as a layer on a map. It can display population by property, by street and by selecting an area.

The application is composed by the core tier and the frontend tier. The core tier consists of a servlet written in Java which loads data from the Publication System. To load the data it queries the SPARQL API and retrieves the data using the Jena library, transforming it into Java Objects. The frontend tier is built with Google Web Toolkit and it just composes the data on the web page that the user can see. The Figure 3 shows an example of the web application.

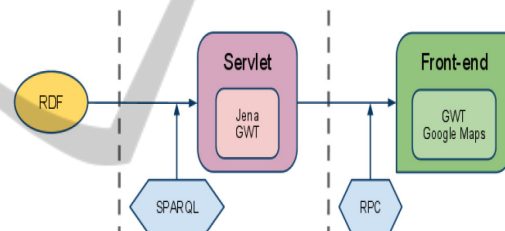


Figure 2: The Reutilization System.

4 CONCLUSIONS

In this work we have presented a real case of OGD with the collaboration of a municipal institution. This example shows that the application not only must have attractive user interfaces but data analytics to transform data to valuable knowledge.

Open Government Data involves a diverse group of technical and non-technical people to be thinking around the strategy. It will take some time before this new interaction of the OGD becomes a wide reality, meanwhile citizens will demand public data to be accessible in different formats and applications.

The key to success of the OGD is that the three stakeholders (citizens, administrations and developers) will complement each other. Citizens must collaborate and offer feedback to administrations and developers to improve the quality of the data and the applications. This impro-

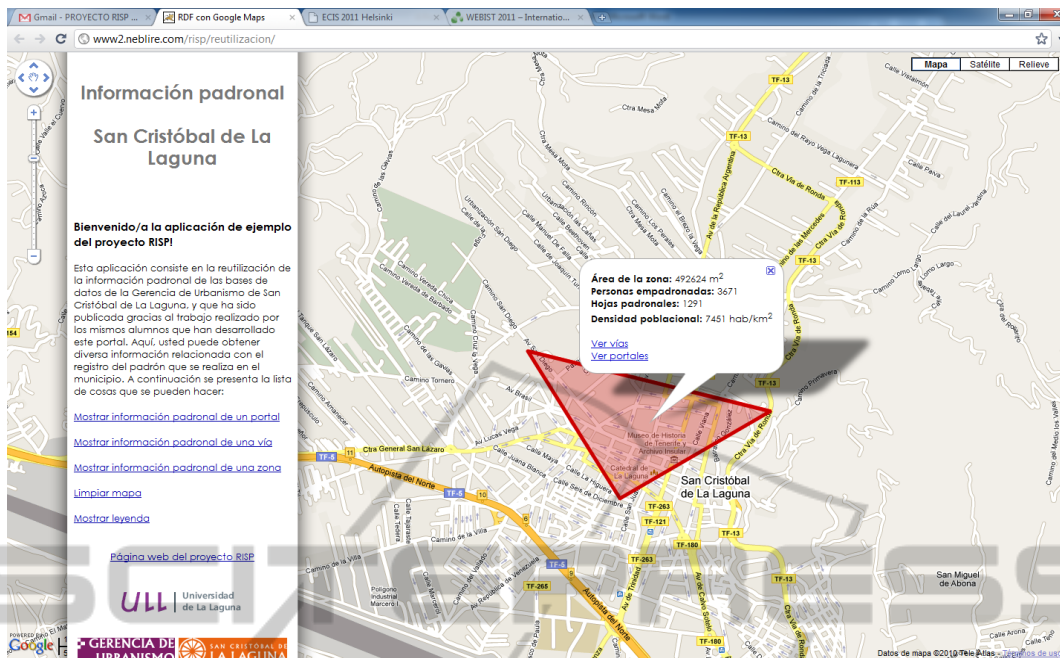


Figure 3: Web application using reutilization of public data.

vement will return to citizens in the form of better web or mobile applications. The public administrations can benefit of the new applications for their decision making at low or non-costs.

Governments are responsible for providing the data in open formats such as RDF, and companies or individual citizens will develop web and mobile applications. However, it is not easy at all to share data to the people because some concerns must be considered (security, private information, trust).

ACKNOWLEDGEMENTS

We would like to thanks to Roberto Martín, Luis López, Vicente Quilis and Javier Cabrera from the Gerencia de Urbanismo, Ayuntamiento de La Laguna, Canary Islands, Tenerife, Spain. Their interest and support have made this project viable.

REFERENCES

Directive 2003/98/EC of the European Parliament and of the Council, of 17 November 2003. *On the re-use of public sector information*. http://www.epsplus.net/content/download/3234/34944/file/English_1_34520031231en00900096.pdf.

O'Reilly, Tim. 2005. *What Is Web 2.0*. <http://oreilly.com/web2/archive/what-is-web-20.html>.

Lathrop, D., Ruma, L.. 2010. *Open Government. Collaboration, transparency and participation in practice*. ISBN 978-0-596-80435-0. O'Reilly Media.

Open Government Working Group. 2007. *Open Government 8 principles* http://resource.org/8_principles.html.

Berners-Lee, Tim. 2010. *Putting Government Data Online* <http://www.w3.org/DesignIssues/GovData.html>

RDF. 2010. *Resource Description Framework (RDF) Model and Syntax Specification*. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>.

OWL. 2010. *Web Ontology Language*. <http://www.w3.org/TR/owl-features/>

O'Reilly, Tim. 2010. *Government as a Platform*. <http://opengovernment.labs.oreilly.com/ch01.html>.