

Open Data as a key factor for developing expert systems: a perspective from Spain

Luz Andrea Rodríguez Rojas¹, Juan Manuel Cueva Lovelle¹, Giovanni Mauricio Tarazona Bermúdez², Carlos Enrique Montenegro²,

¹University of Oviedo, Asturias, Spain

²Francisco José de Caldas District University, Bogotá, Colombia

Abstract — The open data movement is relatively new but very significant, and potentially powerful. The overall intention is to make local, regional and national data available in a form that allows for direct manipulation. This paper is based on analyzing the current context of the Open Data initiative in Spain, from its origins and concepts, the legal framework, current initiatives and challenges that must be addressed for effective reuse of public sector information.

Keywords — Open Data, Interoperability, linked data, e-government.

I. INTRODUCTION

THE new technologies of information and communication technologies (ICT) have dramatically changed the ways of access to public sector information facilitating information processing and its dissemination. The main producer of information are the governments, which, in the exercise of their functions, create, collect, treat, store, distribute and disseminate large amounts of information of various fields.

The e-government, the reuse of public sector information, with open data initiative has led to the spread of numerous campaigns opening public sector data and administrative transparency. Spain has not been unresponsive to these initiatives. However, despite the good position in the world ranking for the development of Open Data initiatives, the efforts made so far are still insufficient.

Artificial intelligence contributes to the advancement of government management models, where issues relating to electronic government and open, ensure transparency of data online and public accountability, and citizen participation.

This paper shows the current situation of open data initiative from a Spanish perspective, its importance, challenges and opportunities.

II. BACKGROUND

In recent years has begun a revolution called "Open Data" caused by the wide array of government data that are significant not only because of the quantity and centrality but also because most government data are public by law [1] and that entails access and reuse of public information [2].

The Open Data initiative is related to the e-government and the web presence administration and is about to publish the Public Sector Information in standard formats, open and interoperable, facilitating their access and enabling reuse.

Many central governments are making government information more easily available on the Web to the public and in formats that citizens can reuse. This occurs in response to the needs of society for transparency, participation, collaboration, innovation, accountability, economic development, job growth, cost reduction, interoperability, among other [3].

The open data value chain including the source which can be public or private, infomediaries which are responsible for data processing and end users who may be enterprises or citizens (Fig. 1).

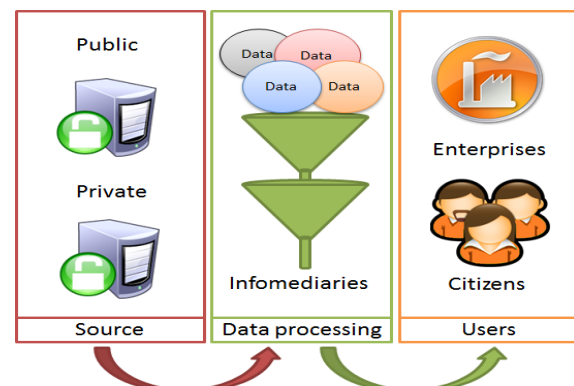


Fig. 1. Open Data value chain

Technical semantic and organizational interoperability difficulties are evident throughout the value chain of open data. From the lack of a clear legal framework and political leadership[4–6] until socio-technical impediments from the perspective of the open data user [7].

Reuse of Public Sector Information (RPSI) is the main objective of Open Data initiatives. RPSI means making public information generated by the public sector available to individuals and business. Although its great economic potential is not yet possible to quantify the benefits of this initiative.

A study, of use of open government data (OGD) from data.gov.uk [8] puts forward five processes of OGD use (fig. 2): Data to fact, information, interface, data or service.

For improving the open data process research should be about the impediments in users groups with the objective to propose strategies allowing for the accomplishment of the desired impact.

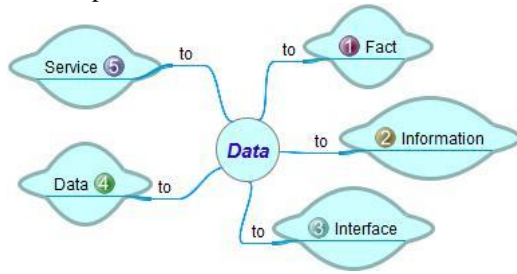


Fig. 2. Five processes of OGD

Moreover, linking information from different sources is a key for further innovation. The ultimate goal of Linked Data is to enable the use of the Web like a single global database [9] with easy accessible information obtained from different sources [10]. If data can be placed in a new context, more and more valuable applications will be generated [11]. It is clear that interoperability and standards are a key to fully benefit from OGD. Therefore, Linked Data is essential to generate an information management through standards to ensure interoperability. This needs to be considered at three levels [12]:

Organizational level: coordinated processes in which different organizations achieve a previously agreed and mutually beneficial goal.

Semantic level: precise meaning of exchanged information which is preserved and understood by all parties, enables organizations to process information from external sources in a meaningful manner.

Technical level: planning of technical issues involved in linking computer systems and services, It includes aspects such as interface specifications, interconnection services, data integration services, data presentation and exchange, etc.

Not only users, but also other stakeholders, may have greater freedom to make choices when the level of interoperability increases, there are other benefits to make interoperable government systems, for example, better data quality, the greatest opportunity in the delivery of information and finally non-tangible benefits such as gaining greater citizen trust[13]. Furthermore using the right architectural model to deploy interoperable e-government services results in a direct, positive impact on Gross domestic product (GDP) [14].

III. OPEN DATA IN WORLD

Open data is closely related to the maturity models as the e-government. The World Bank described e-government as the use of information technologies by government agencies that have the ability to transform relations with citizens, businesses, and other arms of government [15].

Governments around the world are at various stages of e-Government readiness and implementation, European

countries generally taking the top spots [16]. According to the 2012 United Nations E-government Survey rankings, the Republic of Korea is the world leader (0.9283) followed by the Netherlands (0.9125), the United Kingdom (0.8960) and Denmark (0.8889), with the United States, Canada, France, Norway, Singapore and Sweden close behind [17].

Data.gov is an initiative to put hundreds of thousands of public datasets in an easily usable and retrievable form in one location. It means democratizing access to data, it is one of the first initiatives, which since 1999, is driving the U.S. government.

The United Kingdom repository data.gov.uk is a large repository of public information that the British government launched to provide transparency to the management, data and financial issues on the UK government are an example of transparency.

The European Union (EU) recently released the beta version of its open data portal. This web page contains all public information generated by EU institutions. A study conducted ten portals, edited by public bodies member countries of the European Union in 2012 showed that most have less than two years and are in a phase of organization and implementation, so that consistency and data coverage is very uneven [18].

In Latin America, in Chile, there are three open data initiatives: an open data portal which publishes, on a single website, public information sets in more than one format, Chile Library of Congress the first Latin American state institution validated with five stars Open Data by CTIC Foundation and Open Data catalogue by the Chile's Transparency Board. In Brazil, there are open data portal and Federal Senate Legislative Open Data as national initiatives and other regional such as "Minas em Numbers" data, statistics and indicators of mines.

Several countries in Latin America are studying and making experiments with Open Data, the same is happening, on a much smaller scale, in a few parts of Asia and Africa. Kenya is the first developing country to have an open government data portal, the first in sub-Saharan Africa and second on the continent after Morocco[19].

IV. OPEN DATA IN SPAIN

A. legal framework

In 1998, the European Commission launched the "Green Paper on Public Sector Information in the Information Society" highlighting the need to improve synergies between the public and private sectors in the information market. In this Green Paper lies the foundation for the regulation of access to public information, transparency, reuse and trading it. The Green Paper proposals inspired the Directive 2003/98 of The European Parliament and of The Council of 17 November 2003 on the re-use of public sector information.

Meanwhile in Spain the law 37/2007 was passed for the regulation of the legal regime applicable to the reuse of documents produced or held by the Administrations and public

sector bodies and the law 11/2007 which recognizes the right of citizens to interact with the public administrations by electronic means and regulates basic aspects of using information technologies. The Royal Decree 1495/2011 and the Royal Decree 1671, respectively, partially implemented such laws (fig. 3).

Recently, the EU Council approved the amendment of Directive 2003/98/EC which regulates the reuse of Public Sector Information. Among the innovations introduced is expanding its scope, the obligation to publish a standard format, open and processable by automatic means. On the other hand, member states will be held accountable for the implementation of the Directive every 36 months and the Commission should develop guidelines for licensing unique rates and publication preferred datasets.

Spain is working on transparency law and executing projects and initiatives in order to fulfill The European eGovernment Action Plan 2011-2015.

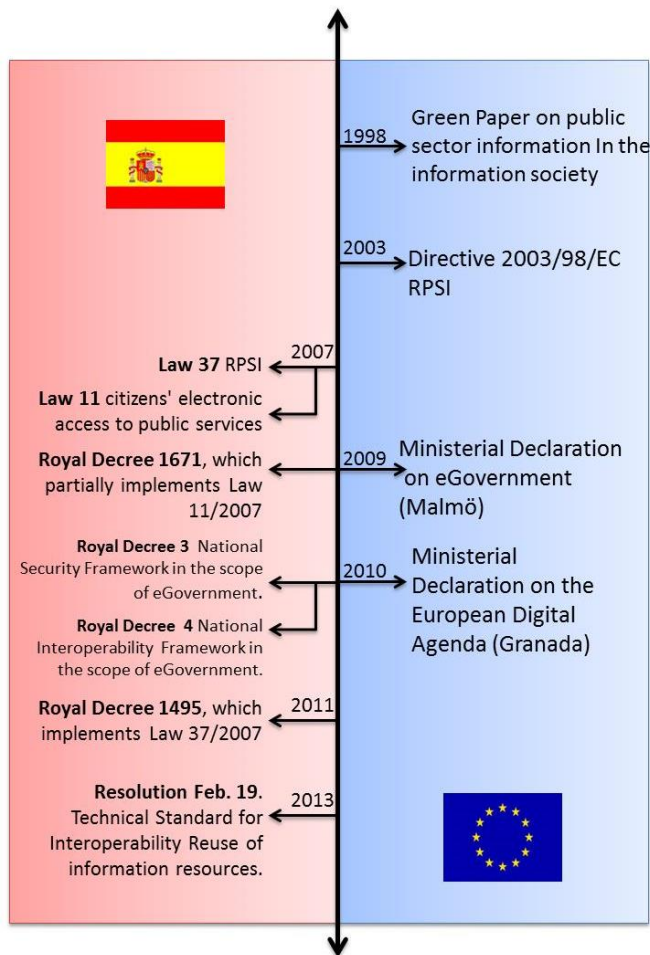


Fig. 3. Legal framework 1998-2013.

B. National and Regional initiatives

In 2005 Spain's government approved through the Avanza Plan, the first national action line dedicated specifically to promote the opening of public sector information. In 2009 it created the project Aporta, fruit Avanza2 Plan. Two of his main achievements were the launch of www.aporta.es and the

publication of the Aporta Guide. The Guide offers a series of recommendations and best practices in open public sector information and is a good starting point for any work related to Open Data. Through the Aporta Program, what has been working to boost the provision of public information by the different administrations, with several initiatives for effective implementation of the culture of reuse of information within the administration.

In March 2010 (fig. 4), the Public Information catalogue was launched, the first single access to available data sources in Spain's public sector. In June 2011, the "Characterization infomediary sector 2011" was published a study, which showed the situation of open data in Spain. In that same year, the initiative datos.gob.es was launched, taking over from project "Contribute" and assumed its commitment to open public sector information. Its main objective is to promote the publication, improve access and encourage reuse of public information corresponding to the General State Administration.

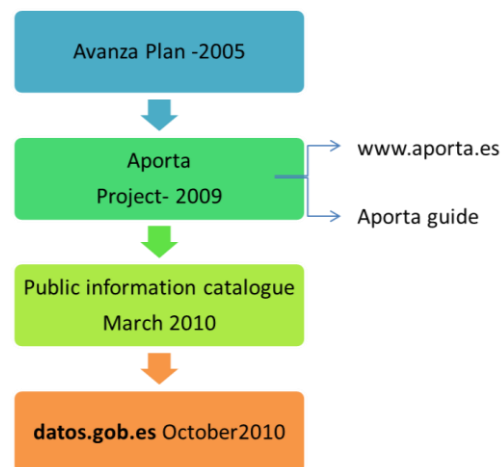


Fig. 4. Initiatives open data in Spain

Other examples of policies reuse of information is opening data from the General Directorate for Cadastre and National Geographic Institute.

The CTIC Foundation (Center for the Development of Information and Communication Technologies in Asturias) is a organization, constituted by a group of firms from the Information and Communication Technologies field, and the Government of the Principality of Asturias. The datos.fundacionctic.org is the CTIC website dedicated to Open Data. Its goal is to help government and public bodies in the publication of your information in accordance with current regulations.

There are also some regional projects that have been developed by several Public administrations to share their data, together with CTIC Foundation. Table I shows the main initiatives undertaken in provinces of Spain.

Although the government is working on the reuse of data and important steps have been taken in recent years, there is still much to be done in the field of open data.

Open data Euskadi was born in 2010 as the first open data

portal in Spain. It has been awarded the "Project Transparency Referent in central, regional or Local Government" in 2013 and shows a genuine opening of public data for reuse. In Asturias on the other hand, despite linked data, the number of published datasets is very low and not updated frequently. Gencat data catalogue uses different formats according to the data type, this variety makes it hard access. Most of the data sets published in the Open Data Junta de Castilla y León are in proprietary formats o Comma Separated Values (CSV). It particularly emphasized the launch of Open data Córdoba in 2011, with an investment of 382,000 [21] euros, which is not currently operating and for which the reasons are unknown.

TABLE I
OPEN DATA REGIONAL INITIATIVES IN SPAIN

Region	Open data initiatives	Catalogue status [20]
Andalucía	Andalucía Open Data Catalog	Open formats
	Open Data Universidad Pablo de Olavide - Sevilla	Open formats
	Open Data Córdoba	Extinct
Asturias	Asturias Public Data	Linked Data
	Gijón Public Data Catalogue	RDF Data
País Vasco	Open Data Euskadi	RDF Data
Navarra	Open Data Navarra	Open formats
	Pamplona Open Data	Open formats
Aragón	Aragon Open Data	RDF Data
	Zaragoza Public Data Catalogue	Linked Data
Galicia	Abert@s (Galicia Open Data)	RDF Data
Castilla La Mancha	Portal de Datos Abiertos de JCCM	RDF Data
Cataluña	Lleida Open Data	Open formats
	Gencat data catalogue	RDF Data
	Open Data Terrassa	RDF Data
	OpenData Sabadell	Open formats
	Badalona Open Data	Open formats
	OpenData Barcelona	RDF Data
	Observatori de la Ciutat Open Data Sant Boi	Open formats
Islas Baleares	Balearic Islands Open Data	RDF Data
Castilla y León	Open Data Junta de Castilla y León	RDF Data

Only 60% of open data catalogs of the initiatives mentioned above announced the update frequency and none makes visible downloaded datasets registration number. Also it is difficult to know the true value of such data. Furthermore Lleida and Sant Boi portals are only available in Catalan.

V. CHALLENGES AND OPPORTUNITIES

The impact of open data initiatives is emerging, while it is true, Spain has consolidated initiatives, a legislative framework to promote the reuse of public information, steady growth in electronic government there are many challenges still to face (fig. 5).

Technically speaking it is necessary to establish common parameters for all the autonomous communities regarding the

information published, quality, format, licensing, processing and loading at each site. Regarding the semantic the challenge is to standardize formats and establish a standard reference vocabulary that various administrations can use to facilitate the linking of data and broaden the scope to nationwide solutions.

Finally, from an organizational perspective strong leadership from General State Administration is required in order to identify what are the data that generate added value, define indicators to measure and evaluate initiatives and recognize the importance of open data as an engine for the generation of economic benefit, transparency and improving vertical and horizontal interoperability in public administrations.

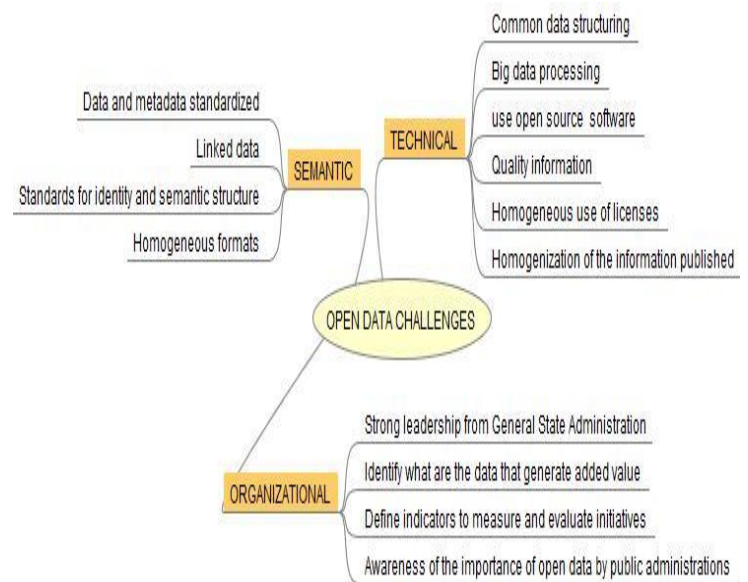


Fig. 5.Challenges open data

The actual citizen participation is essential, the frustration and lack of trust in institutions in many countries are high, so it's no surprise when people express skepticism that opening government data won't help much in fixing things.

VI. CONCLUSIONS

Despite the significant advances in the reuse of public sector information, in Spain there are still things to be done. The progress has been slow if we consider that for over 15 years, the European Commission has been working in this scope.

Probably the most difficult problems to overcome are not technical or semantic, but rather those relating to the understanding that the information belongs to everyone and that the empowerment of the public administrations is essential in order to share its information with others, and to raise awareness of the potential of opening data for reuse.

In this sense, research on new expert systems applied to the generation of an assistant to the search, extraction and understanding of information in large volumes of data is completely innovative, and allows for a tremendous value to projects that leverage the potential of Open Data.

REFERENCES

- [1] Open Knowledge Foundation, "Manual de los Datos Abiertos," *Portal de Datos Abiertos del Gobierno de Misiones*, 2012. [Online]. Available: http://www.datos.misiones.gov.ar/directorio/documentos/Manual_de_Datos_Abiertos.pdf.
- [2] A. Naser and G. Concha, "Datos abiertos: Un nuevo desafío para los gobiernos de la región," *Naciones Unidas*, 2012. [Online]. Available: http://www.eclac.org/publicaciones/xml/7/46167/DatosAbiertos_17_04_2012.pdf.
- [3] Proyecto Aporta, "Guía Aporta sobre Reutilización de la Información del sector Público," 2011.
- [4] N. Huijboom and T. Van Den Broek, "Open data: an international comparison of strategies," *European Journal of ePractice*, pp. 1–13, 2011.
- [5] M. Fioretti, "Open Data : Emerging trends , issues and best practices administration," 2011. [Online]. Available: http://www.lem.sssup.it/WPLem/odos/odos_report_2.pdf.
- [6] E. D. Rico and R. M. Calenti, "Open Data y RISP: generando valor social y económico. Transparencia e innovación en la eAdministración," *Santiago de Compostela: Colexio Profesional de Enxeñaría en Informática de Galicia*, 2011.
- [7] A. Zuiderwijk, M. Janssen, and S. Choenni, "Socio-technical Impediments of Open Data," *Electronic Journal of eGovernment*, vol. 10, no. 2, pp. 156–172, 2013.
- [8] T. Davies, "Open data, democracy and public sector reform. A look at open government data use from data.gov.uk," University of Oxford, 2010.
- [9] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data - The story so far," *International Journal on Semantic Web & Information Systems*, vol. 5, pp. 1–22, 2009.
- [10] J. L. Redondo-García, V. Botón-Fernández, and A. Lozano-Tello, "Linked Data Methodologies for Managing Information about Television Content," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 1, no. 6, p. 36, 2012.
- [11] F. Bauer and M. Kaltenböck, "Linked Open Data: The Essentials," 2011.
- [12] F. B. Vernadat, "Technical, semantic and organizational issues of enterprise interoperability and networking," *Annual Reviews in Control*, vol. 34, no. 1, pp. 139–144, 2010.
- [13] G. Concha and A. Naser, "El desafío hacia el gobierno abierto en la hora de la igualdad," 2012.
- [14] L. Madrid, "The Economic Impact of Interoperability Connected Government," *Microsoft Corporation*, 2012.
- [15] The World Bank, "e-Government," 2011. [Online]. Available: <http://go.worldbank.org/M1JHE0Z280>. [Accessed: 15-Feb-2013].
- [16] W. Olatokun and B. Adebayo, "Assessing E-Government Implementation in Ekiti State, Nigeria," *Journal of Emerging Trends in Computing and ...*, vol. 3, no. 4, pp. 497–505, 2012.
- [17] United Nations Department of Economic and Social Affairs, "E-Government Survey 2012," 2012.
- [18] F. Ramos, R. Arquero, I. Botezán, S. Cobo, A. Sala, R. Sánchez, and F. del V. Gastaminza, "De la reutilización de información del sector público a los portales de datos abiertos en Europa," *BiD: textos universitarios de biblioteconomía i documentació*, vol. 29, 2012.
- [19] "Kenya openData," 2013. [Online]. Available: <https://www.opendata.go.ke/>.
- [20] Fundación CTIC, "Public Dataset Catalogs Faceted Browser," 2013. [Online]. Available: <http://datos.fundacionctic.org/sandbox/catalog/faceted/>.
- [21] Diario Córdoba, "Nace la plataforma 'Open data Córdoba' para buscar empleo o crear empresas," 2011. [Online]. Available: http://www.diariocordoba.com/noticias/cordobalocal/nace-plataforma-open-data-cordoba-para-buscar-empleo-o-crear-empresas_625578.html.



Luz Andrea Rodriguez is a Ph.D. ABD in Computer Engineering from the Oviedo University, Asturias, Spain (2012). She has a Master degree in Engineering and website design (2012), is a Specialist in hygiene and occupational health (2010) and a Industrial Engineer, at the Francisco José de Caldas District University, Bogota, DC, (2008). Her research interests include Web Engineering, eHealth, eGovernment, eBusiness, opendata and Tourism.



Juan Manuel Cueva Lovelle is a Ph. D. from Madrid Polytechnic University, Spain (1990). He is Mining Engineer from Oviedo Mining Engineers Technical School in 1983 (Oviedo University, Spain). From 1985 he is a Professor at the Languages and Computers Systems Area in Oviedo University (Spain). ACM and IEEE voting member. His research interests include Object-Oriented technology, Language Processors, Human-Computer Interface, Web Engineering, Modeling Software with BPM, DSL and MDA.



Giovanni M. Tarazona Bermudez is a Ph.D. ABD in systems and computer services for internet at the Oviedo University, Asturias, Spain (2012). He has a Diploma of advanced studies November 2007 from the Pontifical University of Salamanca. He is a Industrial engineer and works at Francisco José de Caldas District University. He is a founding member of KAIZEN PBT GROUP Ltda. Director Research Group on Electronic Commerce Colombian GICOECOL Director of several research projects. His research interests include eBusiness, eCommerce, DSL and MDA.



Carlos Enrique Montenegro Mari is a Ph.D. in systems and computer services for internet from University of Oviedo, Asturias, Spain (2012). He has a Diploma of advanced studies 2008 of the Pontifical University of Salamanca. He is MSc. Science in Information and Communication Systems from the Francisco José de Caldas District University. He is a System engineer. His research interests include Object-Oriented technology, Language Processors, Modeling Software with, DSL and MDA.