CONSUMPTION OF PUBLIC INSTITUTIONS’ OPENDATA BY SPANISH CITIZENS

Manuel Gértrudix, María-Carmen Gertrudis-Casado and Sergio Álvarez-García

Manuel Gértrudix, PhD in information sciences, is a professor of digital communication, and academic director of the Center for Innovation in Digital Education, at the Universidad Rey Juan Carlos (Spain). Specialist in the area of digital communication, his research analyzes the application of new media including topics such as communication, digital storytelling, e-learning, and competitive intelligence. He has been the technical director of new technologies at the Ministry of Education and Science of Spain. He has been a visiting researcher at the University of Central Florida (USA) and at the University of Stirling (UK).

http://orcid.org/0000-0002-5869-3116
manuel.gertrudix@urjc.es

María-Carmen Gertrudis-Casado holds a PhD in communication sciences from the University Rey Juan Carlos, in Madrid. She is a researcher with the Ciberimaginario group. Since 2005 she has worked as a specialist in the application of information and communication technologies to education for the National Institute of Educational Technologies and Teacher Training, University Rey Juan Carlos, Learning 360 Soluciones, Universidad del Norte (Colombia), and the foundation Enlace Hispano Americano de Salud. Her research focuses on immersive and interactive education and communication, and information and knowledge management.

http://orcid.org/0000-0001-6770-203X
Research Group Ciberimaginario (URJC-Icono14)
carmen.gertrudis@ciberimaginario.es

Sergio Álvarez-García is an associate professor and researcher at the Universidad Rey Juan Carlos, in Madrid. He teaches undergraduate and graduate courses in the field of multimedia communication. He is the content projects coordinator in the Center for Innovation in Digital Education. He holds a PhD in information sciences from the Complutense University of Madrid. His research is related to the application of information and communication technologies to journalism, advertising, communication, and learning.

http://orcid.org/0000-0001-9477-2148
sergio.alvarez@urjc.es

Universidad Rey Juan Carlos, Faculty of Communication
Edif. Departamental, Camino del Molino, s/n. 28943 Fuenlabrada (Madrid), Spain

Abstract

The consumption practices of Spanish citizens of open data published by public institutions are evaluated, within the context of the implementation of the Transparency Portal in Spain. Through an online survey, representative of the studied universe, and through an analysis of descriptive statistics, contingency tables and association measures for nominal variables, the degree of awareness about the availability of open data offered by public bodies are compared, their use is described, and reasons for their use are identified. The results reveal low penetration of the open data culture, and citizen distrust
of the actions carried out by public institutions for the access, use, and reuse of open data, which will require a change of orientation in communication and stimulation policies.

Keywords
Open data; Public institutions; Transparency; Empowerment; Digital citizenship.

Resumen
En el contexto de la puesta en marcha del Portal de la transparencia en España, se evalúan las prácticas de consumo de datos abiertos publicados por las instituciones públicas, que realizan los ciudadanos españoles. A través de una encuesta online, representativa del universo de estudio, y mediante el análisis de estadísticos descriptivos, tablas de contingencia y medidas de asociación para variables nominales, se contrasta el grado de conocimiento sobre la disponibilidad de datos abiertos que ofrecen los organismos públicos, se describe el uso que hacen de éstos, y se identifican las razones que les motivan a utilizarlos. Las principales conclusiones revelan una baja penetración de la cultura de datos abiertos, y cierta desconfianza de la ciudadanía sobre las acciones que realizan las instituciones públicas para favorecer el acceso, uso y reutilización de los datos abiertos, lo que exige un cambio de orientación en las políticas de comunicación y dinamización.

Palabras clave
Datos abiertos; Organismos públicos; Transparencia; Empoderamiento; Ciudadanía Digital.

1. Introduction

The implementation of the Transparency Portal in Spain is the end result of a lengthy legal battle aimed at providing access to public information. It gives a practical and enforceable specification to initiatives about access to open data from the public sector, thereby facilitating Government, citizen participation, and co-innovation (Cobo, 2012).

Open data may be considered an “industrial revolution of data”

The purpose of this process is to facilitate citizen access to open data, which is generated by public institutions to carry out their activities. Data that are accessible by any citizen for use, treatment, or re-use within the limits defined by the legislation, are presented in the open format in a direct and standardized way.

However, a culture of transparency and accountability requires multiple approaches to create an open Government; specifically, an improvement in citizen participation and a more involved civil society (Gertrudis-Casado; Gértrudix; Álvarez-García, 2016a).

To better understand this it is essential to discover what the consumption practices of Spanish citizens over 16 years of age are, in relation to open data published by public bodies, based on the citizen’s level of knowledge and use, and the mode of access.

1.1. The Transparency portal

In accordance with Law 19/2013 (España, 2013), the Transparency portal was launched on the 10th of December 2014, with over 59,000 visits in the first eight hours of its operation (EFE, 2014).

http://transparencia.gob.es

The portal gave access to 530,000 records related to Central administration and was organized into three areas:
- organizational information
- legally significant information, and
- economic information.

According to information published by the Transparency portal, on August 31st 2016, the number of requests was 4,275, of which 93.54% corresponded to completed files, 6.22% to files being processed, and 0.23% to files held in administrative silence. In terms of the resolution of requests, 68.77% had been resolved with the request being authorized. As for the rest of the requests, in accordance with Law 19/2013, 23.93% were not accepted (article 18), 3.23% were denied (article 14 and 15), and 4.08% were dismissed. 89.61% of requests were received through electronic access with a password, the rest on paper. Finally, in terms of the subject matter of requests, they were mostly related to
- statistical data;
- budgets/expenses/revenues;
- salaries and payments to personnel; and
- legislation (processing, acts, and reports).

The Transparency Information Unit of the Ministry of Finance and Public Administration receiving the most requests (Ministerio de la Presidencia, 2016).

1.2. Reuse of public sector information

Open data has been considered a kind of “industrial revolution of data” (Hellerstein, 2008). Linked with open government, it offers opportunities in which public sector information and its reuse become an asset for citizen empowerment and is recognized as a basic resource for the evolution of businesses that reuse said information and bring it added va-
In 2009 in the United States, transparency was defined for the first time in the *Memorandum on transparency and open government* (Obama, 2009). Transparency, participation and collaboration were established as pillars of open government, and the first international initiative was launched:

https://www.data.gov

Shortly afterwards, directories for New Zealand, Australia, and the United Kingdom were implemented (the last one with the Tim Berners-Lee support).

https://data.gov.uk
http://data.gov.au
https://data.govt.nz

The *Malmö Declaration*, in November 2009, established the aim of

“citizens and companies having to be empowered by electronic administration services [...] as well as by increasing access to public information, greater transparency and more effective methods for the involvement of participants” (Ministerio de la Presidencia, 2009, p. 2).

In 2011, the *Open Government Partnership* was created, initially with eight member countries, which Spain later joined in April 2012. Currently sixty-four countries collaborate on this international platform to facilitate governments in being


Thus, new relationships have developed between public institutions and citizens: e-administration, e-democracy, e-government, e-governance (Eudel, 2011), based on transparency and access to information as a requirement for governance (Prats-Català, 2005). It implies the need for a strategic government in the public sector, capable of articulating these relationships (Muñoz, 2005) as interdependence-association-coproduction/co-responsibility between the government and the private and social associations (Aguiar-Villanueva, 2007).

The practice of governance has been supplemented by a new form of managing resources, *new public management* (García-Sánchez, 2007), which implies the incorporation of business criteria, allowing for valuable results to be obtained (accountability and value for money) by applying economic, efficiency, and effectiveness principles in public management. It has inherited the principles of total quality management (TQM) (Summers, 2006).

In November 2005 Spain started opening public sector information, and its promotion was entrusted to the public business entity Red.es. Two years later, *Law 37/2007* (España, 2007) was implemented on the 16th of November, and the reuse of public sector information was approved, implementing *Directive 2003/98/EC* in national legislation,
and setting out the basic principles for reuse. In 2009 the Aporta Project was established by the Spanish Ministry of Finance and Public Administration (Minhap) and the Ministry of Industry, Energy, and Tourism (Minetur). From this initiative, in 2010, the first Public Sector Information Catalog was created, accessed through the datos.gob.es portal. This catalog was developed from the directory of open access databases, sponsored by the administration of the State and created by Group Publidoc-UCM (2008).

"Aporta has established annual meetings to evaluate the impact of open data from public organizations"

Subsequently, the Aporta annual meetings (Encuentros Aporta), which began in 2014, established an agenda to evaluate the impact that open data has for both the citizen and institutional intelligence policies. It has been a long process (Minhap; Minetur, 2015, p. 4), and the results will be seen in the future, which will require the tendering of a large cross-section of society to extend the term of re-users to "citizens, developers, sociologists and journalists, among other groups" (Minetur; Minhap, 2015, p. 5).

In the Digital agenda for Spain, the course of action that the plan develops for encouraging the digital economy and digital content incorporated in Axis III, a program for the reuse of public sector information, seeks to "increase the publication of data sets by Public Administrations and increase business activity in the infomediary sector" (Minetur; Minhap, 2015).

On the 11th of July 2015, Law 18/2015 (España, 2015) of the 9th of July came into force, modifying Law 37/2007 of the 16th of November (España, 2013), on the reuse of public sector information. This law fulfills the European Union mandate of correspondence with Directive 2013/37/EU and updates, in accordance with active legislation, the conditions for reuse of public sector information, especially in the regulation of formats, wherever possible promoting its availability in open, machine-readable formats, along with their metadata.

1.3. Social and economic effects of opening data

In 2013, according to National Observatory for Telecommunications and the Information Society (Ontsi), there were over 500 companies dedicated to the reuse of public sector information and over 4,000 employees linked to said sector, producing products and/or services from the reuse of public data, with a turnover of between 450 and 500 million euros (Ontsi, 2015a).

In terms of the infomediary sector, according to the Ontsi study, there is a trend toward non-exclusive use of sources, using data coming from the public and private sectors to generate products and services for the market studies sector, the economic and financial sector, and others, such as support companies, database generators or price comparison creators (Ontsi, 2015b). It is also increasing the reuse of data in smart city portals through apps (Abella; Ortiz-de-Urbina-Criado; De-Pablos-Heredero, 2015).

Some projects, such as Meloda, propose a model for measuring the reuse of open data based on four criteria: technical standards, form of access, legal framework, and data model (Abella; Ortiz-de-Urbina-Criado; De-Pablos-Heredero, 2014).

In terms of homogenization and standardization, the work of the World Wide Consortium (W3C) in Spain is notable, through the Open Data Group and RISP (Reuse of Information in the Public Sector).

Likewise, the UNE 178301 (Aenor, 2015) regulation, created by the specific group of smartcities, defines a set of indicators divided into five pillars (political, organizational, technical, legal, and economic) and the measurements necessary to quantify the level of open data in Spanish cities. The regulation defines 10 sets of data, and its vocabulary aims to serve as a reference for governments to publish their data in a standardized, interoperable, and ultimately open way.

"There are over 500 companies dedicated to the reuse of public sector information with over 4,000 employees"

2. Methodology

2.1. Aims and hypothesis

The general aim of the research is to examine the consumption practices of Spanish citizens of open data published by public bodies. To do so:

- the extent of knowledge and use by citizens, of open data offered by public bodies, is compared;
- the media diet of individuals who use open data is identified;
- the sociodemographic profile of individuals who use open data is established;
- the reasons for their use are identified; and
- the value of public bodies opening data are revealed, as well as the quantity of open data currently available, the ease of access to them, and the ease of using, analyzing, and interpreting them.

The hypotheses are:

- The lacking culture of transparency and open data in Spain is directly related to the low levels of awareness and usage of open data by citizens.
- The citizens who are most aware and most often use open data are those who have a more diverse media diet with less focus on mass media.

2.2. Methodological strategies

To obtain information from citizens, a computer assisted online questionnaire was carried out, using the LimeSurvey system at a national level (including the Canary Islands, and the cities Ceuta and Melilla). The study universe is the internet-using Spanish population over 16 years of age.
The structured questionnaire had 16 questions with an estimated completion time of 12-15 minutes (Gértrudix; Gertrudis-Casado; Álvarez-García, 2016b).

The sample size was 1,183 interviews, with a random sample stratified by Autonomous Community (Spanish administrative regions) with proportional allocation, carried out by using the email address directory of the company Toluna USA, which is statistically representative and proportional to the Spanish population by size of home and province. The sampling error, for a 95.5% confidence level (two sigmas), and \( P = Q \), is ±3.16% for the whole sample and in the case of a simple random sample. The field work was carried out from the 6th to the 17th of October 2014.

In the data analysis, descriptive statistics were used to evaluate the profile, use, and evaluation of the open data, contingency tables to estimate whether the level of knowledge and use of the open data published by public bodies may be conditioned by the sociodemographic variables studied, and to check the degree of association between the variables a Person Chi-squared test was carried out, as well as other measures of association for nominal variables (based on the Chi-squared test) such as Phi, Pearson’s C Contingency Coefficient, and Cramer’s V Contingency Coefficient.

3. Results

3.1. Profile and use of open data

To evaluate the profile and use of open data, the awareness and use, and reasons for use were analyzed, and subjects were classified using cluster analysis.

Knowledge and use of open data published by public bodies

The number of individuals who stated that they used public administration open data was 232, representing 20.6% of the sample. This figure rises, however, when they were asked whether they know said open data, making up 40% of the sample. From the data obtained, a profile of open public institution open data users was identified, taking into account the most revealing characteristics on sample distribution, from sociodemographic variables (Gértrudix; Gertrudis-Casado; Álvarez-García, 2016c).

Reasons for using open data published by public bodies

The main reasons for using open data published by public bodies was related to personal context, studies (22.5%), personal curiosity (22.3%), and personal issues (20.3%) being the main reasons, followed by actions in daily life (10.6%). It is notable that social motivations obtained very low percentages: protests and complaints (8.4%), social demands (8.1%), common causes (6.3%), and collective actions and social movements (1.5%).

Classification through clusters

In order to obtain a classification, in terms of the frequency of use of open data and their level of compliance with the type of dissemination carried out by public institutions on open data, a cluster analysis was carried out with two of the variables from the study:

- “Frequency of use of open data from public institutions”
- “Sources of open data from public institutions”, being the only variables that allowed different groups of respondents to be distinguished (Gértrudix; Gertrudis-Casado; Álvarez-García, 2016c).

Two types of individuals with different characteristics clearly appeared:

- Those individuals who are less linked with open data from public institutions: they rarely used the information based on open data from public institutions at all levels; and disagreed that public institutions provided sufficient open...
data on their activities, that these were offered in an accessible and easy to find manner, or offered citizens simple tools and applications to analyze and interpret the data provided.

- Those individuals who are more linked with open data from public institutions: defined by greater use of information from public institutions at all levels, notably using information that the European Union, Spanish government, autonomous communities, and municipalities disseminate on the internet about their activities. They had an intermediate evaluation of sources of open data. Despite this, they were clearly more positive toward open data than the first individuals.

Once this cluster classification of the respondents had been obtained, a comparison was made with the variable “Access or use of open data published by public bodies”, showing interesting differences between both groups. The majority of the first group (51.5%) were unaware of the existence of open data, and the percentage of individuals who were aware of it had low usage (around 11%). However, a large proportion of the second group (46.8%) stated that they were aware of open data although they had never used them, and there was a rather high percentage of individuals who were aware of and had used open data (almost 30%).

Most respondents believe that public institutions do not offer enough open data about their activities.

3.2. Evaluation of open data characteristics

There is not a distinct stance on whether public institutions publicize themselves on the internet through their open portals and banks, as 35% neither agree nor disagree. The other evaluations are mainly distributed in two positions on the scale: agree and disagree. Finally, it is notable that there are few cases placed in the most extreme categories.

Most respondents (55.8%) believed that public institutions do not offer sufficient open data about their activities.

Adding to this idea, most of the sample (57%) also disagreed that public institutions provided open data in an accessible and easy to find manner. The majority (50%) also believed that public institutions do not offer easy tools and applications in order to analyze and interpret the open data that they provide.

3.3. Level of awareness and use of open data published by public bodies depending on sociodemographic variables

For this analysis, three groups depending on the level of knowledge and use of open data made available by public institutions are considered: those who have used them; those who know their existence but have not used them; and those who do not know their existence.

Considering the gender variable, there were no great differences in the level of use of open data. A variable which presented a larger difference was the degree of awareness of this type of product, in this case women (58.3%) had lower awareness. In terms of the degree of association between both variables, the Chi-squared ($\chi^2$) statistic value is 11.568 (> theoretical value of 5.99) and its level of significance (<0.001) which indicated that the null hypothesis of independence of variables may be rejected, and it can be established that the variables are statistically associated. However, the measures based on $\chi^2$ (Phi, Pearson’s C Contingency Coefficient, and Cramer’s V Contingency Coefficient) take a value of 0.101 with a critical level of significance of 0.003, indicating that association between the variables was slight.

By age groups, it was observed that there was a greater awareness and use of open data by young people (between 16 and 34 years of age), and only 14.7% stated that they were unaware of it. As ages increased, the percentage of individuals who used open data decreases up to the 35 to 44 year-old age group. In terms of the group that was aware of it, but did not use it, there were not great differences between age groups, except at the extremes, where the youngest were most aware (23.3%) and the oldest were least aware (16.2%). Finally, in terms of those who were unaware of the existence of open data, it is notable...
that it was the youngest group that had the highest percentage (21.6%).

Finally, the Chi-squared value ($\chi^2$=25.426 > $\chi^2$= 15.51) and its level of significance (0.001) indicated that the variables were associated. However, the degree of association was not very high.

As there are few cases in many of the educational stages which preceded university, this variable was divided into two large groups: pre-university and university. It was observed that there was a larger percentage of individuals with a university education who were aware of and had used open data, compared with those who were not aware of it, where there were more pre-university individuals.

The users who are most aware of and most often use open data are those who have a more diverse media diet

However, the association of these variables was statistically significant. The Chi-squared value ($\chi^2$) exceeds the theoretical level for two levels of freedom (16.699 > 5.99 $\chi^2$) and, therefore, the null hypothesis of independence of variables can be rejected and it can be concluded that the level of study and access to open data are statistically related (Gertrudis-Casado; Gertrudis-Casado; Álvarez-García, 2016c).

In the group of individuals who work, the percentage of individuals who were aware of and had used open data was greater (59.5%) than those who were studying (24.1%). As far as their level of association, the Chi-squared statistic ($\chi^2$=41,852) allows it to be concluded that access and use of open data and the labor situation are statistically linked.

In terms of level of awareness and use of open data published by public bodies, in terms of political ideology, there were not great differences. Furthermore, the statistics indicated that variables were not statistically associated.

Despite the statistics indicating that there was an association between the variables, there were not major differences between the three groups mentioned earlier in this section 3.3 in terms of personal economic circumstances, or civil status. Furthermore, both the value shown by statistics and their level of significance (>0.005) indicated that these variables were independent.

3.4. Access or use of open data published by public bodies by sources used to search for information on the internet

One of the two hypotheses for this study stated that the users who are most aware of and most often use open data are those who have a more diverse media diet and are less focused on mass media. In order to test this hypothesis a comparative analysis was carried out between the dependent variable “Access or use of open data published by public bodies” and the independent variable “Sources for searching for information on the internet”.

The percentage distributions indicate that the group of those who had accessed and used open data from public institutions were notable for their use of digital communication and public institutions’ and tertiary sector websites, and they made greater use of social networks and news aggregators. Likewise, despite the percentage being small, among those seeking information through alternative citizen journalism information channels and media, there was a greater percentage among those who used open data than those who did not use it.

However, the high value of the Pearson Chi-squared statistic (304,680) and its level of significance (< 0.001) indicate that the variables are statistically linked (Gertrudis; Gertrudis-Casado; Álvarez-García, 2016c).

4. Conclusions and discussion

Despite the importance of expanding the culture of open data (García-Sánchez, 2007; Marcos-Martín, Soriano-Maldonado, 2011; European Parliament and Council, 2003), there is still low penetration in Spain, and the number of individuals in our study that stated that they had used open data from public institutions was only 20.6% of the sample.

Add to this low level of awareness a distrust by citizens towards public institutions that encourage access, use, and reuse of open data, and the long term effects demonstrated in the latest Aporta meeting are revealed (Minhap; Minetur, 2015; Minetur; Minhap, 2015). A significant number of citizens do not believe that public institutions offer sufficient information about their activities, or that they are easily accessible, or that they offer easy tools and applications to analyze and interpret open data. This may be explained partly by the lack of awareness shown in studies such as the CIS barometer (2012), but also by the lack of continuity which the administrations’ evaluation of these issues has had.

It is important to expand the culture of open data in Spain, because its penetration is currently low

This may be considered as part of the institutions’ lack of will to be transparent and apply the Law of transparency (AIE; CLD, 2015).

This shows the long path that open data still has ahead of it and the need for administrations to be aware that they must implement initiatives which is especially relevant when considering that Law 19/2013 established, in its additional seventh provision, that to improve the impact and effectiveness of its application it would develop “an informative campaign aimed at citizens”.

Spain currently has an abundance of data with zero marginal cost (Rifkin, 2014, p. 333), but the benefits of the data have not reached citizens, possibly because citizens have not been considered (Minhap; Minetur, 2015; Minetur; Minhap, 2015).
Depending on the frequency of use and degree of acceptance of sources of open data we found there were two groups with different characteristics who possessed differentiated levels of awareness, behaviors, and attitudes relating to the phenomenon of open data. There is a relationship between interest, awareness, use, and trust which reveals the importance of public institutions maintaining a proactive attitude toward the design and implementation of initiatives and actions which give value to transparency and the data ope-ning process, as the Open Data Charter (GB, 2013) and the Estudio de la demanda y uso de gobierno abierto en España (2013) from OnTsI indicate.

Contrary to expectations, social causes (protests and complaints, social demands, common causes, etc.) did not motivate the use of open data, as none of these reached 10%. It was instead personal motivations which were the main reasons for use, which is consistent with the results of studies such as Gurstein (2011), Bates (2012), Janssen, Charalabidis and Zuiderwijk (2012), Tinati et al. (2012) and Zuiderwijk and Marijn (2014).

The analysis of the relationship between level of awareness and use of open data published by public bodies and the sociodemographic variables show slight associations in terms of gender and age variables, although there was a greater awareness of open data in the case of men and a greater use in the age group between 16 and 34 years of age, falling as age increases. The level of education did have a statistically significant association and the higher the level of studies then the greater the awareness and use of open data, with the greatest number being individuals who had completed university studies. This is also the case with the occupation variable. Individuals who work were more likely to be aware of and use open data provided by public institutions. Significant relationships were not found in terms of personal economic circumstances, civil status, or geographic location by autonomous community in Spain.

The hypothesis that users who are more aware and make greater use of open data are those who have a more diverse media diet and are less focused on mass media was considered and showed that both variables are statistically related. It is therefore shown that access to data portals and the options which these provide help citizens to expand their media diet (Hasebrink; Domeyer, 2012; (García-García; Gértrudix; Gértrudix, 2014).

Acknowledgement

This work was supported by the Ministry of Economy and Competitiveness of Spain under grant “OpenDataCitizen. Digital citizenship and open data access: Citizen empowerment through social media in the digital environment” (CSO2012-30756), and Grupo de Excelencia investigadora URJC-Banco de Santander Anima+D-Ciberimaginario.

5. References


España. Minhap; Minetur (2015). Informe conclusiones me-
Consumption of public institutions' open data by Spanish citizens


G8 (2013). Open data charter. Lough Erne. https://goo.gl/1hDj9a


Tinati, Ramine; Carr, Les; Halford, Susan; Pope, Catherine (2012). “Exploring the impact of adopting open data in the UK government”. In: Digital futures 2012, Aberdeen, GB. http://eprints.soton.ac.uk/id/eprint/344808