UACJ INVESTIGACIÓN

Título del Proyecto de Investigación a que corresponde el Reporte Técnico:

An Analysis of the Supply of Open Government Data

Tipo de financiamiento

Sin financiamiento

Autores del reporte técnico:

Benito Alan Ponce Rodríguez
Raúl Alberto Ponce Rodríguez

An Analysis of the Supply of Open Government Data

Resumen del reporte técnico en español

Un índice de la publicación de datos de gobierno abierto publicado en 2016 por la Open Knowledge Foundation muestra que existe una variabilidad significativa en la oferta del país de este bien público. ¿Qué explica estas diferencias entre países? Adoptando un enfoque interdisciplinario basado en la ciencia de datos y la teoría económica, desarrollamos el siguiente flujo de trabajo de investigación. Primero, recopilamos, limpiamos y fusionamos diferentes conjuntos de datos publicados por instituciones como The Open Knowledge Foundation, Banco Mundial, Naciones Unidas, Foro Económico Mundial, Transparencia Internacional, Economist Intelligence Unit y la Unión Internacional de Telecomunicaciones. Luego, llevamos a cabo la extracción de características y la selección de variables basadas en el conocimiento del dominio económico. A continuación, realizamos varios modelos de regresión lineal, probando si las diferencias entre países en el suministro de datos gubernamentales abiertos pueden explicarse por diferencias en las estructuras económicas, sociales e institucionales del país. Nuestro análisis proporciona evidencia de que las libertades civiles del país, la transparencia del gobierno, la calidad de la democracia, la eficiencia de la intervención del gobierno, las economías de escala en la provisión de bienes públicos y el tamaño de la economía son estadísticamente significativas para explicar las diferencias entre países en la oferta. de datos gubernamentales abiertos. Nuestro análisis también sugiere que la participación política, las características sociodemográficas, las variables ficticias demográficas y de distribución del ingreso global no ayudan a explicar el suministro de datos gubernamentales abiertos del país. En resumen, mostramos que las diferencias entre países en la gobernanza, las instituciones sociales y el tamaño de la economía pueden explicar la distribución global de los datos gubernamentales abiertos.

Palabras clave: Ciencia de los datos; datos gubernamentales abiertos; gobernabilidad e instituciones sociales; determinantes económicos de los datos abiertos.

Resumen del reporte técnico en inglés

An index of the release of open government data published in 2016 by the Open Knowledge Foundation shows that there is significant variability in the country's supply of this public good. What explains these cross-country differences? Adopting an interdisciplinary approach based on data science and economic theory we developed the following research workflow. First, we gather, clean, and merge different datasets released by institutions such as The Open Knowledge Foundation, World Bank, United Nations, World Economic Forum, Transparency International, Economist Intelligence Unit, and International Telecommunication Union. Then, we conduct feature extraction and variable selection founded on economic domain knowledge. Next, we perform several linear regression models, testing whether cross-country differences in the supply of open government data can be explained by differences in the country's economic, social, and institutional structures. Our analysis provides evidence that the country's civil liberties, government transparency, quality of democracy, efficiency of government intervention, economies of scale in the provision of public goods and the size of the economy are statistically significant to explain the cross-country differences in the supply of open government data. Our analysis also suggests that political participation, sociodemographic characteristics, demographic and global income distribution dummies do not help to explain the country's supply of open government data. In summary, we show that cross-country differences in governance, social institutions and the size of the economy can explain the global distribution of open government data.

Keywords: data science; open government data; governance and social institutions; economic determinants of open data.

Usuarios potenciales

Estudiantes, académicos, investigadores independientes o grupos inter o multidisciplinarios cuyo interés es analizar el tema de los datos abiertos bajo una perspectiva cuantitativa y aplicando técnicas de ciencia de datos y econometría.

Reconocimientos

Este proyecto de investigación tiene una perspectiva interdisciplinaria en la cuales colaboraron investigadores del Instituto de Ingeniera y Tecnología (IIT) y el Instituto de Ciencias Sociales y Administración (ICSA)

1. Introduction

Open data (OD) refers to information that has been generated by public or private entities and then it is published under a license that allows its use, reuse, and distribution freely [1]. Information collected and released from the public sector (i.e., transportation, pollution, agriculture, education, health, census, among others) is referred to as Open Government Data (OGD) [2] . The public sector is considered one of the main contributors to the open data movement due to the vast amount of information that generates [3]. According to [4] during the last years, there has been an increase in the number of countries that are adopting open data policies as part of their governmental agenda. Authors also argue that this trend is related to the potential benefits that OGD offers as a shared value (social and economic). From the social perspective, OGD is considered as a trigger of transparency, accountability, fight against corruption and empowerment of citizens. The economic aspect of OGD is related to foster innovation, enterprise opportunities and job creation because OGD is considered as a production asset in the digital economy.

Additional evidence of the global interest in the open data topic is the recent creation of different portals in which governments consolidate their data from different public entities (i.e., education, health, transportation) on a single website in order to release their data for free and collective use. Some examples of these portals developed by governments are the US[1], Canada[2], Brazil[3], Mexico[4], or the European Data Portal[5] funded by the European Commission. Other aspects related to the open data interests are the initiatives constituted in conjunction by citizens, academics, and non-governmental organizations, that are creating indexes such as the Global Open Data Index (GODI)[6], Open Data Barometer (ODB)[7], Open Data Watch (ODW)[8], Open Data Impact (ODI)[9] which are measuring the amount of data published by different governments around the world as well as potential benefits and challenges (technical, legal, economic, social) that these public datasets (i.e. education, health, transportation) are generating in society.

Although there has been an increased interest in the phenomenon of open government data, most research has been conducted applying qualitative methodology through surveys, case studies, and desk research focusing on diverse topics such as challenges and barriers in adopting and implementing open government data initiatives and other qualitative studies have been focused on the release, provision, or value of these public datasets [5–7]. However, there is a gap in the literature analysing and measuring the determinants of the supply of open government data adopting a quantitative approach. This work pretends to fill this gap and contribute to the state of the art of open government data providing a statistical analysis explaining countries' variability of the release of open government data through economic, social, and institutional factors. According to the Global Open Data Index published in 2016 by the Open Knowledge Foundation (OKF)[10], there are significant differences across countries in the supply of open government data. In particular, Australia, the United Kingdom, and France obtain the highest GODI scores reported by the Open Knowledge Foundation (meaning that these countries contribute the most to the supply of open government data) while countries such as Myanmar, Barbados, Malawi, Botswana obtained the lowest records on the GODI score (meaning that, in a global comparison, these countries contribute the least to the supply of open government data). This leads us to the following question: *What explains the high heterogeneity in the global supply of open government data?*

The objective of this paper is to provide an answer to this question by extending a single academic perspective due to this research is based on an interdisciplinary approach aligned by the fields of data science and economics. The intersection point of these disciplines lies in analysing and estimating the

[1] https://www.data.gov/open-gov/
[2] https://open.canada.ca/en/open-data
[3] http://www.dados.gov.br/
[4] https://datos.gob.mx/
[5] https://www.europeandataportal.eu/en
[6] https://index.okfn.org/
[7] https://opendatabarometer.org/?_year=2017&indicator=ODB
[8] https://opendatawatch.com/
[9] https://odimpact.org/
[10] https://okfn.org/

determinants of the heterogeneity in the supply of global open government data by means of gathering information from different sources, featuring extraction and variable selection, modelling through the implementation of statistical methods, and explaining the effect and relationship of this heterogeneity. On the one hand, the data science approach is implemented in order to systematically create a data pipeline collected from different portals. This task is executed following an OSEMN process that stands by Obtaining, Scrubbing, Exploring, Modelling and iNterpreting the information collected from several sources. Then, we apply feature engineering in order to extract and analyse by using a regression model that seeks to analyse the statistical association between some political and economic determinants of open government data. To estimate our model of regression analysis, we develop a sample with country-cross section data with data of the Global Open Data Index (GODI) for the year 2016. In this process we solve empirical issues that arise in the regression analysis such as multicollinearity, heteroscedasticity, missing data, outliers, and high dimensionality with our target variable (open government data).

On the other hand, economic theory is adopted to develop an empirical analysis (using our data pipeline) for the analysis of variables and their justification based on domain knowledge. Open government data is considered as a pure public good [8]; that is to say, we consider open government data satisfies two important properties: it is a non-excludable (once open government data is provided then any person, who seeks access, can have access to that good) and it is a non-rival good (the consumption of open government data by some agent does not preclude the consumption of the same good by everyone else). Applying this theoretical framework, we test if political and social institutions such as civil rights, transparency, quality of democracy and political participation, as well as economic and sociodemographic characteristics at the country level (such as the size of the economy, the efficiency of the government, the demand for internet services, the median age of the population of a country and the size of population) can explain the global variability in the supply of open government data.

Using a cross country regression model our analysis provides evidence that cross country differences in governance and social institutions such as civil liberties, government transparency and the quality of democracy are statistically significant predictors of cross-country differences in the supply of open government data. Our estimates suggest that the government's transparency and civil liberties have a marginal positive and statistically significant effect on the supply of open government data. In our model, our variable that captures changes in the demand of web resources, that is the penetration of users (the proportion of internet users over the country's population) is also positively and statistically significant in all of our estimated models.

In addition, our indicators of the efficiency of government intervention and economies of scale in the provision of public goods (analysed through the variable population in each country) are also statistically significant predictors of cross-country differences in open government data. Our models also provide weak support to the hypothesis that open government data is a normal good: that is to say, countries with higher income are associated with higher levels of supply of open government data. Finally, our estimates suggest that political participation, sociodemographic characteristics of citizens, demographic dummy variables and dummy variables capturing the global distribution of income do not help to explain cross country differences of the supply of open government data. In summary, we find evidence that cross country differences in the supply of open government data are associated with the heterogeneity of social and political institutions, and economic factors are also correlated with the supply of open government data.

It is relevant to mention that the main limitation of our analysis is that we use cross section data for our regression analysis which limits the generality of our results. We decided to use data from the GODI index for the year 2016 because this is the most up to date data on GODI. Even if there is data for the Global Open Data Index for other years, the Open Knowledge Foundation has clearly stated that changes in methodology in the calculation of the GODI index make unsuitable the comparison of data between years 2016 and other years. This limits the study of what factors could explain the changes of GODI over time. However, this limitation could be eased as long as more data sets become available in the future that allow other forms of regression analysis such as regression with panel data that might improve the properties of estimation and hypothesis testing as well as the generality of the results.

The structure of the paper is as follows: section two includes a brief literature review postulating the technical, social, economic, and political determinants of global open government data. Section three describes the data collection, the preparation process for our analysis and the identification of the linear regression model. Section four contains the results of our analysis. Section five concludes.

2. Literature Review

The adoption and implementation of open data is a socio-technological phenomenon that has been studied by different disciplines trying to understand and estimate its dimensions and barriers [9–11] . For instance, the technical outlook is associated with the relevance of improving data interoperability, quality, accessibility, usability, accuracy, platforms, and infrastructure needed in order to release open data [12–17]. The social stance refers to the empowerment that data offers to society. For example, the potential benefits that the information released by governments could produce through transparency and accountability on citizens [18–21]. The economic point of view is related to possible impacts on the economy that open data could offer through the creation of new business, products, and services as well as employment [22–24] . This perspective also includes the crucial role that innovation plays as a driver of economic growth in the private and public sectors using open data [25–30] . The political perspective covers the strategies, policies, and impacts of the data released by the state [31–34]. The data published and freely accessible by public entities is referred to as Open Government Data (OGD). This particular kind of data plays an important role in the open data movement because it is considered as one of their main supporters through legislations such as the Open Data Directive[11] or global political initiatives like the Open Government Partnership (OGP)[12]. These political actions aim at increasing efficiency, promoting transparency, empowering citizens and driving a knowledge-based economy through the release of data generated by public sectors. [35] claims that the creation of open data policies is essential for defining the financial and technological infrastructure required, publication process definition, legal framework certainty, and political sustainability of open government data (OGD). The author also argues that open data policies should disseminate the economic and social value of OGD in order to stimulate the use and reuse of it in society. [36] argue that the release of OGD is relevant because there are datasets collected by different sectors and for specific purposes (i.e., transportation, pollution, agriculture, education, health, census). The authors also claim that OGD is a driver for innovation and business opportunities for society. Finally, they argue that the infrastructure of these data sets is paid by taxpayers; therefore, this information is considered a public good.

2.1 Determinants of the Supply of Open Government Data

In this section we develop an analysis of the determinants of the supply of open government data. Hence, we explain the incentives of policy makers in government to provide goods and services. As we mentioned before, in this paper we consider open government data as a pure public good which has two important properties: it is a non-excludable (once a pure public good is provided then any person, who seeks to have access, can have access to that good) and it is a non-rival good (the consumption of the good by some agent does not preclude the consumption of the same good by everyone else). In our analysis, we consider that households and firms demand open government data because they find value in it. That is to say, households and firms might find open government data as valuable because this information might help them to make rational and informed decisions. This information can also be used to foster their objectives such as engaging in civic activities, political debates, and other activities regarded as desirable for the case of households and in the case of firms open government data might help them to make more efficient decisions (see [22,37,38]).

The literature on public economics has made important contributions to the study of the provision of this type of goods and this literature can be classified in two distinctive lines of research. The first line is the normative theory and the second is the positive theory of the provision of public goods. The normative literature has emphasized that the preferences of households for private and public goods, the technology of production, and the costs of taxation that finance these goods are the main determinants of the provision of public goods (for a comprehensive review of the normative literature on public goods see [39] and more recently [40]).

In contrast, the positive literature on public goods has emphasized that, in addition of household's preferences and technology of production, governments are suppliers of public goods and candidates to public office are elected through a democratic process. It should be pointed out, the supply of open data has a production cost, and therefore governments need to allocate public budgets for the collection and administration of data. This means, that the allocation of budgets that allow the supply of open government data is subjected to a regular process of political negotiation in congress and the executive power. Therefore, the provision of public goods could be explained by electoral incentives: political candidates seek to win public office and they compete for votes in an election to form the government and make

[11] https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information

[12] https://www.opengovpartnership.org/

decisions over public policy, see [41] and [42]. For this reason, politicians have incentives to provide public goods that benefit a significant proportion of voters in the electorate with the hope of attracting votes in the election and maintaining political support while politicians hold office.

To be more specific about how electoral competition creates incentives for politicians to provide different levels of goods and services, we describe in detail the quid-pro-quo of models of electoral competition (see [41] and [42]). In a democracy, voters with different socio-demographic characteristics such as age, gender, income, etc. might demand certain goods and services from the government because they benefit from these goods and services. Hence, candidates might consider that the distribution of demands of voters for goods and services from the government might be characterized by figure 1. For purposes of exposition, we assume *g* as the size of the provision of the public good, hence figure 1 shows that there is might be voters who would like the lowest size of the public good equal to *gmin* (maybe because he or she does not benefit from the provision of this good), while *gmax* is the size of the provision of other voters who wants the highest level of *g* in the distribution (maybe because the personal characteristics of these voters make them to benefit a great deal from this good). Every point in the line shown in figure 1 represents the ideal policy demanded by a certain voter and the position *gMV* is the ideal policy demanded by the median voter, that is to say, the voter who is in the middle of the distribution of policies demanded by all voters participating in the election.

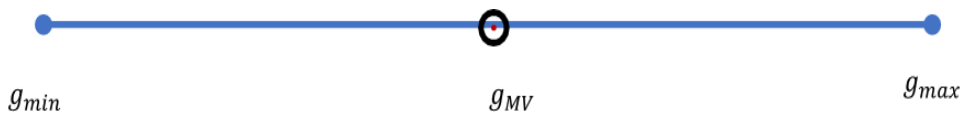

Figure 1. Distribution of Policies Demanded by Voters.

Models of electoral competition consider that a voter will vote for the party that provides the policy that is closer to the voter's own preferences over policies. To see this, assume two parties, say parties 1 and 2, competing for the vote of a particular voter with ideal policy given by *gh* (shown in figure 2). This voter will vote for party 1 because the policy position of this party is closest to the voter's own preferences for this public good or service.

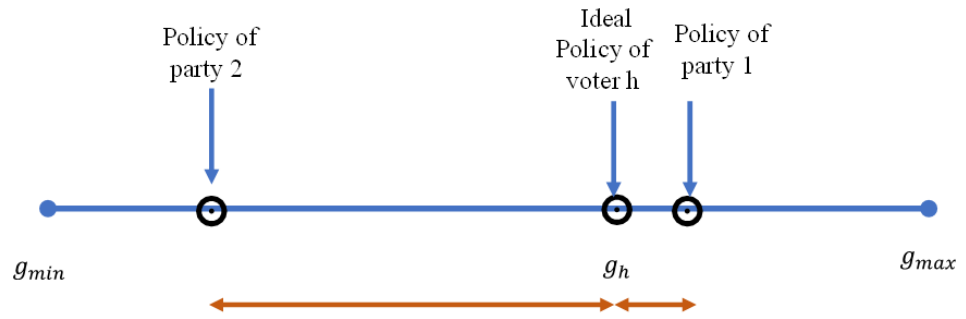

Figure 2. Policy Positions of Parties and the Choice of the Vote

Hence, models of electoral competition predict that parties who want to maximize the expected votes to be received in the election should decide to offer the provision of the public good and service demanded (or desired) by the median voter (see figure 3). That is, the government should select a level of its policy equal to *g=gMV*; By so doing, the expected proportion of the vote for each party is 50% of the vote. If any party deviates from providing the median voter policy, then the party expects to receive a proportion of the vote lower than 50% of participating voters and will lose the election. Hence, models of electoral competition make a strong prediction that can be tested empirically: parties who want to maximize the electoral support from voters in the election should estimate the demand for goods and services of the median voter.

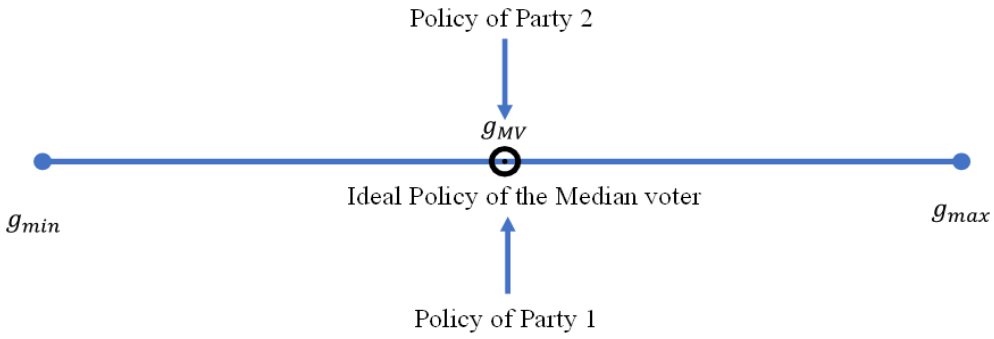

Figure 3. Prediction of Models of Electoral Competition.

A further example of how this mechanism works is the following: assume the demand of the median voter for goods and services from the government increases (perhaps because the median voter has more income and desires more government services) then policy makers in the government should increase the supply of government services to satisfy the demand of the median voter. It is relevant to mention that there is a great deal of evidence suggesting that public goods, such as education, health, and infrastructure (roads and bridges) that are provided by national and subnational governments are correlated with the incentives of elections and political competition. For global empirical evidence of such relationship covering 118 countries for three decades see [43].

The positive literature on public goods has emphasized that if elections matters and if there is perfect electoral competition (as a similar concept to the idea of perfect economic competition) then parties select the ideal provision of public goods of the median voter [42]. In this case there is electoral accountability meaning that elections create incentives for the provision of public goods that satisfy at least a majority of voters. However, if there is imperfect electoral competition and parties have preferences for public goods (that is to say, individuals controlling parties desire certain types and levels of public goods) then parties might select the ideal policy of activists inside parties or the ideal policy over public goods of a minority group of voters in the electorate, (see [44])[13] . In this latter case, there might be little electoral accountability and the provision of public goods might be different relative to the ideal policy of the median voter in the economy (which might be considered as the ideal public policy for the society as a whole). Hence, the quality of the democratic process matters to determine the degree of electoral accountability and the size of the provision of public goods. Hence, for democracies with electoral accountability, if there is an increase in the demand of voters for open government data, well-functioning governments might increase the supply of open government data to satisfy the demand of voters for that public good, see [45–48].

The demand for public goods can also be related with political participation. Citizens express their demand for public goods and services through voting in elections, see [40], [42] and [49]. The political participation of citizens can be observed by different political channels such as voting in elections, attending meetings to express their demands for specific goods and services to elected representatives in congress and the executive power, or by contributing to political campaigns. Hence, we expect that more political participation lead to more accountability and better governance in democracies. Therefore, in democracies in which there is electoral accountability, higher political participation should lead to a better match between the public goods and services demanded by citizens and the supply of such services by the government.

[13] In an election, there could be imperfect electoral competition when a party does not have strong incentives to use its policy positions to attract a majority of the votes in the election. This could be the case if voters do not vote for parties based on their policy positions but instead on party identification (whether a voter self-identifies with a party), or the choice of the vote could be strongly determined by other non-policy issues such as the personal characteristics of candidates (age, gender, etc.). For instance, if a significant proportion of voters, (for purposes of exposition, let's say 30% of voters), vote for some party based on party identification then this party does not necessarily select the policy of the median voter, because this party only needs another 21% of the vote to win the election. In this case, the policy positions of parties might be heavily influenced by the preferences over policy of candidates or influential groups of voters inside of the party. Hence, there might be low electoral accountability and if there is an increase in the demand of voters for open government data, the government might not respond by changing the supply of open government data. In this case, the demand of voters for that public good might not be satisfied.

A well-functioning democracy is also related with civil liberties of citizens and the provision of public goods, (for analysis along these lines see [50]) . Civil liberties are associated with the access of a free printed and electronic media which provides relevant information to all citizens. Civil liberties can also be related with freedom of association and protest, and more relevant to our analysis, with political institutions that foster the free access to the Internet. Hence, we expect that more civil liberties are positively associated with less political restrictions to access the Internet and therefore more demand of content freely available on the Internet. In this line of thinking, the supply of open government data should also be positively related to transparency from government. Transparency might help well informed voters and economic agents to make rational decisions about the functioning of the government. Hence, voters might demand that their government provides useful information about the decisions of public policy of their elected officials. Therefore, in countries in which citizens demand more transparency we could expect that their government satisfy this demand by providing more open government data.

The literature has also recognized that the sociodemographic characteristics of individuals, such as age, gender, marital status, etc. might be important determinants of the demand of public goods and services (for a classical analysis on this issue see [51] and for a literature review of the impact of socio-demographic characteristics on the size of government spending on public and other type of goods and services see [50]). Hence, changes in sociodemographic characteristics of households are related with changes in the demand for public goods and services (for instance a change in the average age of individuals might lead to a change in the demand of certain services such as public education, public welfare, etc.). Therefore, changes in sociodemographic characteristics of voters in a democracy might lead to changes in their demand for public goods and governments have incentives to change their supply of public goods accordingly.

Most theoretical models that seek to explain the demand of private and pure public goods consider whether public goods are normal, neutral or inferior, see [52] and [53] . If a good is normal, then an increase in income of households leads to an increase in the demand for such good. If a good is inferior, then an increase of the household's income leads to a fall in the demand of such good (when income increases households might substitute the demand of low-quality goods for high quality goods which might explain why the demand of certain goods might fall as the household's income increases). A neutral good does not respond to changes in the household's income see [54] . Hence, we could expect that an increase in the country's income might lead to an increase (fall) in the demand of open government data if this good is normal or an inferior good and governments might respond by increasing (reducing) the supply of open government data.

In addition, most theoretical models that study the provision of pure public goods consider the size of population as an important determinant of the provision of pure public goods, see [39], [52] and [53]. As we mentioned before, a pure public good is non-excludable (once a pure public good is provided then any person can have access to that good) and non-rival (the consumption of the good by some agent does not preclude the consumption of the same good by everyone else). Under these circumstances, the non-excludable property of a pure public good means that there could be economies of scale in the costs of providing a pure public good (see [55]). This means that the per-capita costs of providing a pure public good are decreasing as the cost of public goods are shared among more people. In addition, an increase in the size of population might also be associated with an increase in the size of the tax base that finances the provision of a pure public good (see [52] and [53]). This, in turn, leads to a fall in the per-capita cost of providing public goods which increases the demand for this type of goods. Therefore, we could expect that an increase in the size of population of a country might lead to an increase of the demand of open government data and governments might respond by increasing the corresponding supply of such goods.

In summary, to guide the empirical analysis to be conducted in the following sections we have relied on formal economic theory to characterize empirically verifiable tests on probabilistic determinants on the provision of open government data. In our analysis, we consider open government data as a pure public good because it satisfies two properties identified in the economic literature: that is to say, the non-excludable and non-rival good properties. Based on the contributions of economic theory, we state several hypotheses about a probabilistic relationship between the supply of open government data by country and the quality of democracy, the country's political participation, civil liberties and transparency, the sociodemographic characteristics of the country, the size of the economy, the size of the country's population, and the demand of content freely available on the Internet.

3. Material and Methods

3.1.- Data Collection and Pre-processing

A common task in economics and data science fields is the collection of datasets from different sources in order to discover knowledge, patterns and trends. This activity represents some challenges such as the diversity of the data structure, formats, time consistency, among others. In this research, we adopted the OSMEN workflow methodology, which is the acronym of Obtain, Scrub, Explore, Model, iNnterpret proposed by [56] to deal with these challenges. This methodology is proposed in the data science research community in order to systematically collect data, provide research transparency and results reproducibility.

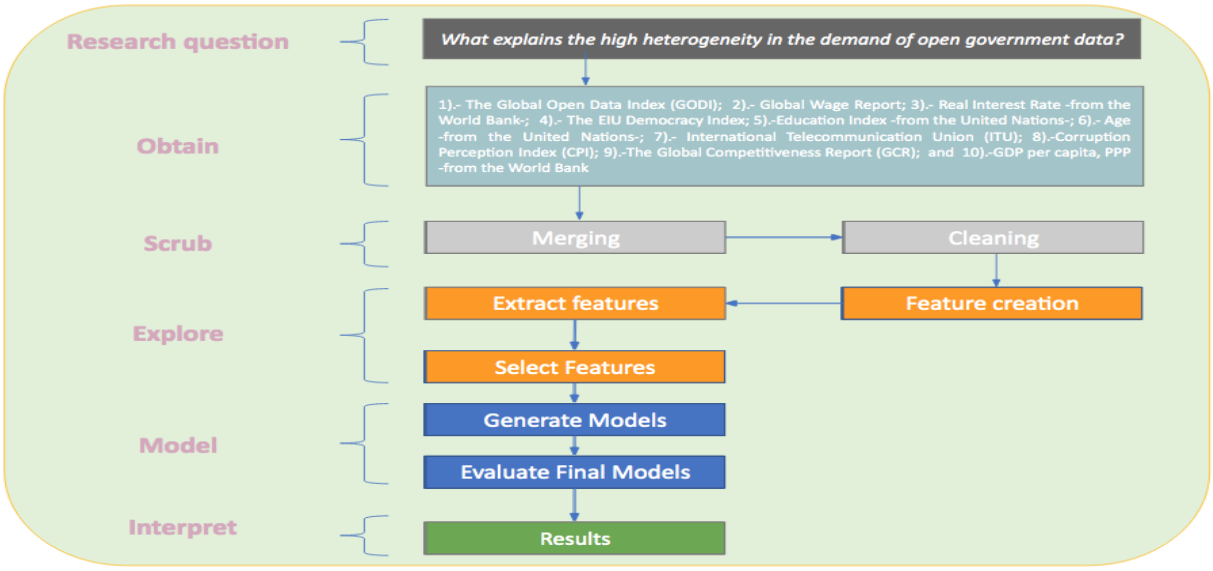

Figure 4. illustrates the data science pipeline developed for our research.

Following this workflow to solve our research question, the first step is to select and gather the data from several sources such as 1).- The global open data index (GODI) - from the Open Knowledge Foundation-; 2).- Global wage report -from the International Labor Organization-; 3).- Real interest rate -from the World Bank-; 4).- The democracy index - from the Economist Intelligence Unit-; 5).-Education index -from the United Nations-; 6).- Age -from the United Nations-; 7).- Internet use -from the International Telecommunication Union; 8).-Corruption perception index (CPI) -from Transparency International; 9).-The global competitiveness report (GCR) -from the World Economic Forum-; and 10).-Gross domestic product (GDP) per capita, PPP -from the World Bank-. In this research, we define the GODI indicator as our dependent variable (also called the response or target variable) and the other information extracted from these datasets are our independent variables (also called labels variables).

Once the phase of data acquisition is completed, the next tasks are data integration and cleaning. The former involves analysing and merging heterogeneous datasets. For example, some information is published in a long format and other datasets as wide formats and containing different periods of time. The latter is associated with keeping consistency among datasets. For instance, homologating the name of the countries because some datasets have different names labels (e.g., in some datasets the country name is denoted such as the United States of America, or Venezuela, RB, and in other datasets, the country name appear as the United States or Venezuela respectively). Another aspect related to the cleaning process is to identify elements such as missing values, outliers or other noise elements that can affect the quality of our model [57].

The next step is related to feature engineering, in particular feature creation, extraction, and selection. Some of our collected variables are categorical data; therefore, we need to create new variables in order to perform our models. This process is known as one-hot encoding in machine learning or dummy variables in econometrics. Then, we perform feature extraction implementing principal component analysis (PCA) which is the process of dimensionality reduction from a large number of attributes without losing meaningful information by removing redundant data. This reduction process helps to identify features that could be more conducive to our analysis [58] . After performing PCA, our analysis indicates that there is multicollinearity in our independent variables. The topic of multicollinearity is an ongoing research area in feature engineering due to its implications using diverse datasets [59–61]. For this reason, we include in the next section a variable selection process and robustness check based on an

economic domain knowledge approach [62] and justified on the theoretical background introduced in the literature [63] . In order to complete a full sample with the desired variables for our empirical analysis, our final cross-sectional dataset is constituted by 18 variables and 49 observations during the year 2016. In the next section, we describe our model generation, interpretation, and results.

3.2.- Empirical Analysis

In this section, we test if political and social institutions such as civil rights, transparency, quality of democracy and political participation, as well as economic and sociodemographic characteristics at the country level (such as the size of the economy, the efficiency of the government, the demand for internet services, the median age of the population of a country and the size of population) can explain the global variability in the supply of open government data. To test our hypotheses, we use one of the most popular tools in data science and economics: a linear regression analysis which allows us to estimate the marginal effect of how changes in independent variables (such as the size of the economy of a country, the sociodemographic characteristics of individuals in a country, civil liberties, transparency, etc.) affect the supply of open government data. In the next model (see equation 1) we postulate that cross-country differences in the supply of open government data are associated with political and economic factors that affect the quality of democracy and the incentives of governments to provide open government data:

$$Od_i = \alpha + \beta' X + \varepsilon_i \qquad (1)$$

In equation (1), the differences in the supply of open government data across countries Od_i for $i = 1, \ldots . I$, where the sub-index distinguishes the different countries in our sample, is explained by a set of *k* independent variables contained in the vector X (such as political participation, the size of the economy, socio-demographic characteristics of households, indicators of demand for internet services, etc.). The vector $\beta' = [\beta_1, \beta_2 \ldots \ldots \ldots \beta_k\ ]$ represents the marginal effect of exogenous changes in X_i in our indicator of the supply of open government data, that is to say, $\frac{\partial Od_i}{\partial X_i} = \beta_i$. Our model also allows to test whether the marginal effect of X_i on open government data are statistically significant (or not). Finally, ε_i is a random error term from our model. To be more specific the model we test in our analysis is specified as follows:

$$Od_i = \alpha + \beta_1 GdpPPP_i + \beta_2 Dem_i + \beta_3 PolPar_i + \beta_4\ Liberty_i + \beta_5 Trans_i + \beta_6 Pop_i + \beta_7 Age_i + \beta_8 GovEfficiency_i + \beta_9 IntPen_i + \varepsilon_i \qquad (2)$$

Hence, in (2), our model tests the postulated determinants of the supply of open government data discussed in section 2.1 of this paper. Therefore, we are interested in testing whether the supply of open government data is associated with changes in the economic size of the country (see $GdpPPP_i$ which is the purchasing power parity of the gross domestic product of a country which affects the demand of open government data and its corresponding supply). The effect of the quality of democracy of a country (defined as Dem_i), which is a metric that measures the function, state, and trust of political freedoms and civil liberties through pillars such as political participation of citizens (defined as $PolPar_i$), civil liberties (defined as $Liberty_i$), the transparency of the country's government (see $Trans_i$).

Other determinants in our model are the size of population (defined as Pop_i), and the sociodemographic characteristics in a country (characterized by the median age of the population of a country, see Age_i). Besides, the efficiency of the government (labelled as$GovEfficiency_i$)which is an index that measures and compare per country the burden of government regulation, legal framework performance, and transparency of public policies and our indicator of demand for internet services (defined by $IntPen_i$ or internet penetration) which is the number of internet users as a proportion of the population in each country. To estimate our model of regression analysis, we develop a sample with country-cross section data. Our variable for open government data is the global open government data index (defined as Od_i) and published by the Open Knowledge Foundation (OKF) which provides cross-country differences on the supply of open government data.

To estimate the model in (2) we use a cross section analysis with data on GODI index for year 2016. A well identified regression analysis needs to consider the possibility of endogeneity which might bias the estimates of the model. Endogeneity might arise when changes in the independent variables X´s might be correlated with changes in the dependent variable Y, and changes in Y might also lead to changes in the X's variables. In this case, the marginal effects in the regression model in (2) would not be properly identified. A standard way to solve this issue is to use the independent variables X´s lagged for one period (for technical analysis on this issue see [64]). Hence, to avoid the possibility of endogeneity in our estimates, we use data for the year 2015 for our control variables, that is, we used the lagged

observation for the control variables such as the size of the country, the quality of democracy, political participation, civil liberties, transparency, the size of population, the sociodemographic characteristics in a country, efficiency of the government, etc. In this case, changes in X´s could be correlated with changes in the dependent variable Y but not the opposite case.

In summary, our assessment criteria for our empirical analysis is constituted as follows: first, we use theoretical analysis from the literature on economics on the main determinants of public goods and services provided by governments to identify control variables of the regression analysis (as a way to determine the structure of the X's variables in the regression analysis, see our section 2.1). The theoretical analysis provides a rationale for a probabilistic link between the independent variable (GODI) and the explanatory variables of the model in equation 2. Second, we use standard techniques of regression analysis to determine the best way to obtain unbiased and efficient estimators of the marginal effects of the independent variables over the dependent variable GODI. Third, we conduct a robustness test of our estimates and hypothesis testing by estimating several models (see models I, II, III, IV and V in table 1 in the following section) to test whether our results are sensitive to specific forms of linear regression analysis.

4. Results

We estimate our model with a set of different independent variables for a robustness check of our analysis. Our estimation technique uses ordinary least squares with heteroscedasticity-consistent standard errors [65–68]. This technique allows having credible estimates of the probability distribution functions of the marginal effects β_i associated with the independent variable X_i and, therefore, credible hypothesis testing. Table 1 shows our empirical results, columns (I) through (V) consider different econometric specifications: Model I uses our basic set of explanatory variables described in equation (2). Model II incorporates an interaction term between democracy and internet penetration which allows us to test if the quality of democracy leads to a differentiated response of countries to an increase in the demand of internet services. Model III expands model II by incorporating geographical dummies, model IV incorporates dummies related with the world´s distribution of income and model V incorporates geographical and global income distribution dummies.

Our estimates show that cross-country differences in governance and social institutions such as civil liberties, the quality of democracy and the degree of government transparency are statistically significant predictors of cross-country differences in the supply of open government data. To see this, note that all models I through V show that the marginal effect of liberty on the supply of open government data is positive and statistically significant (at different levels of p value, see table 1)[14]. In addition, the government´s transparency also has a marginal positive and statistically significant effect on the supply of open government data in models I, II, and III, while the coefficient of the quality of democracy is statistically significant in models II, III, IV and V (see table 1), and the interaction term between democracy and penetration of users of internet is positive and statistically significant in all models in which we consider this interaction term.

In addition, our variable that captures changes in the demand of web resources, that is, in our model the variable of penetration of users (that is the proportion of internet users over the country's population) is positively and statistically significant in all of our estimated models. In models II, III, IV and V we include an interaction term to test whether differences in the quality of democracy leads to different responses of governments to changes in the demand of use of the internet., that is Dem*Penetration. The marginal effect of the interaction term is positive and statistically significant in all of our models in which we use this variable. That is to say, all countries in the sample have a marginal positive response in the supply of open government data when they observe increases in the demand of internet services but countries with higher quality of democracy supply more open government data than countries with weaker democracies. This result confirms that governance is an important determinant of the cross differences in the supply of open government data.

However, in our models, political participation and the sociodemographic characteristic of citizens (in our models, the average age of citizens) are not statistically significant in any of the estimated models.

[14] When we refer to marginal effects of a change of one variable and the variable of GODI, this should be interpreted as a probabilistic marginal effect that the increase of one variable is correlated with increases or reductions (depending on the sign of the coefficient) of the GODI index. Hence, our analysis does not show causality but a probabilistic correlation.

The marginal effect of political participation has a positive sign (as expected) while the average age of citizens has a negative sign (as intuition might suggest) in models I, II and III while a positive sign in models IV and V. In addition, our models provide, at best, weak support to the hypothesis that open government data is a normal good: that is to say, countries with higher income are associated with higher levels of supply of open government data, since the positive income effect on open government data is significant only in model (I). Once we include demographic dummy variables and dummy variables of the global distribution of income, the marginal effect of the size of the economy on the supply of open government data is positive but not statistically significant (see models II, III, IV and V).

In our analysis two variables are associated to the efficiency of government intervention (this is the variable *efficiency* in table 1) and economies of scale in the provision of public goods analysed through the variable population. Our estimates show that the efficiency of government of the country has a negative and statistically significant marginal effect on the supply of open government data in all of our models. One possible explanation of this outcome is that an increase in the efficiency of government might lead to more resources available to be spent by governments. However, those available resources are spent on other governmental programs that might have a high electoral impact relative to the choice of supplying more open government data. Therefore, more efficiency of government increases spending in some programs (with high electoral impact) and reduces spending in other programs (with relatively low electoral impact). In addition, the size of population which is considered as a variable associated with economies of scale in the provision of public goods has the expected sign, that is there is a positive and statistically significant marginal effect of population on the supply of open government data in all of our estimated models. As we mentioned before, the non-excludable property of open government data means that there could be economies of scale in the costs of providing this pure public good. This means that the per-capita costs of providing a pure public good are decreasing as the cost of public goods are shared among more people (that is, as the size of population in the country increases). This, in turn, leads to a fall in the per-capita cost of providing public goods which increases the demand for this type of goods, hence governments respond by increasing the corresponding supply of such good.

Table 1. Results of OLS Estimators with Heteroscedasticity-Consistent Standard Errors**.**

Variable	upply of Open overnment data GODI Score (I)	upply of Open overnment data GODI Score (II)	upply of Open overnment data GODI Score (III)	upply of Open overnment data GODI Score (IV)	upply of Open overnment data GODI Score (V)
C	11.04	50.17	55.5180	53.3683	55.9864
Gdppp	0.0002*	0.0002	0.0002	0.0002	0.0002
	(1.6998)	(1.5876)	(1.6380)	(1.3166)	(1.4044)
Liberty	0.4111**	0.4974***	0.5252**	0.4306*	0.4679*
	2.1968	2.5259	2.4246	1.7019	1.6761
Dem	-0.4726	-1.2298*	-1.3926*	-1.4452*	-1.4923*
	-1.1856	-1.8287	-1.6874	-1.8960	-1.6831
Polpar	0.1581	0.1925	0.1978	0.2730	0.2410
	0.7461	0.8889	0.8105	1.1275	0.8072
Age	-0.3381	-0.3287	-0.2821	0.0522	0.0645
	-0.9542	-0.9782	-0.7732	0.1132	0.1178
Transparency	15.8476*	17.48105**	16.1369*	13.9994	14.4925
	1.7610	1.9572	1.7938	1.5910	1.6067

Efficiency	-19.6248**	-22.12**	-20.8876**	-18.6995*	-19.6564*
	-2.0313	-2.2669	-2.1968	-1.8295	-1.94
Population	2.5681***	3.1520***	3.3581***	3.3492***	3.3951***
	3.4713	3.50	3.3188	3.0274	2.8543
Internet-Penetration	0.5345***	-0.0142	-0.1968	-0.2213	-0.2840
	2.9659	-0.0474	-0.4547	-0.4952	-0.5341
Dem*Internet-Penetration		0.0088*	0.0114*	0.0116*	0.0129*
		1.8122	1.8913	1.8565	1.8966
High income			-0.015		-2.8034
			-0.0009		-0.1298
Upper Middle Income			4.8105		1.6383
			0.3667		0.1074
Lower Middle Income			0.6255		-0.0153
			0.0842		-0.0016
ast Asia Pacific				3.9374	2.9713
				0.3864	0.2726
Europe Central Asia				-1.2709	-2.5741
				-0.1179	-0.2231
Latin America Caribbean				8.0636	5.0072
				0.6578	0.3605
Middle East North Africa				0.8853	2.6890
				0.0724	0.2168
North America				2.0636	0.7462
				0.1751	0.0595
Adjusted R-squared	0.6387	0.6605	0.6689	0.6792	0.6824
F-statistic	7.66***	7.39***	5.43***	4.65***	3.58***
Sample	49	49	49	49	49

***P <0.01, **P <0.05, *P < 0.10. All tests are two-tailed. t tests are in below of the corresponding estimates (numbers in parenthesis correspond to the t-test).

Our models, through the individual t-test (the t tests are displayed in table 1 in parenthesis), also suggest that demographic dummy variables and dummy variables capturing the global distribution of income are not statistically significant to explain the cross-country differences in the supply on open government data in our sample (see models III, IV and V). However, the F statistic shows that, jointly, all independent variables considered in models (I) through (V) are statistically significant to explain the cross-country differences in the supply of open government data in our sample. To see this, we test if $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \ldots \beta_k = 0$ (that is, if the joint effect of our independent variables help to explain cross-country differences in the supply of open government data) and the corresponding F statistic shows that we reject the null hypothesis shown above. Therefore models (I) through (V), as a whole, are statistically significant (see table 1).

5. Conclusions

We develop a cross-section analysis to provide tests for institutional, political and economic determinants of cross-country differences in the supply of open government data. We consider open government data as a pure public good and therefore it satisfies two properties: open government data is a non-excludable good (once open government data is provided then any person, who seeks to have access, can have access to that good) and it is a non-rival good (the consumption of the good by some agent does not preclude the consumption of the same good by everyone else). We use an index of the supply of open government data and estimate a cross-section regression model to analyse the cross-country differences in civil rights, transparency, quality of government, the size of the economy, the size of population, political participation, and sociodemographic characteristics can explain cross country differences in the supply of open government data. We also conduct a robustness check of our analysis by estimating five different models of regression analysis that also include dummy variables associated with the geographic heterogeneity and the global distribution of income.

Our analysis provides evidence that cross-country differences in governance and social institutions such as civil liberties, government transparency and the quality of democracy are statistically significant predictors of cross-country differences in the supply of open government data. Our estimates suggest that civil rights and the transparency of government in each country have a marginal positive and statistically significant effect on the supply of open government data. In addition, our variable that captures changes in the demand of web resources, that is penetration of users (the proportion of internet users over the country´s population) is positively and statistically significant in all of our estimated models. Our analysis also suggests a differentiated response of government to changes in the demand of web resources: in particular, if there is an increase in the demand of internet services, countries with higher quality of democracy supply more open government data than countries with weaker democracies. This result also shows that the level of governance in each country is an important determinant of the cross-country differences in the supply of open government data.

In our analysis we include two variables that are associated with the efficiency of government intervention and with economies of scale in the provision of public goods. In this paper we find evidence that the efficiency of government and the economies of scale can also explain the cross-country differences in open government data. The efficiency of government has a negative marginal effect on open government data. One possible explanation of this outcome is that an increase in the efficiency of government might lead to more resources available to be spent by governments. However, those available resources might be spent on other governmental programs that might have a higher electoral impact relative to the choice of supplying more open government data, thus, explaining the negative marginal effect of efficiency on the supply of open government data. In addition, the size of population in a country is considered as a variable associated with economies of scale in the provision of public goods. Our analysis shows that there is a positive and statistically significant marginal effect of population on the supply of open government data in all of our estimated models. As we mentioned before, the non-excludable property of open government data means that there could be economies of scale in the costs of providing open government data. This means that the per-capita costs of providing a pure public good are decreasing as the cost of public goods are shared among more people. This, in turn, leads to a fall in the per-capita cost of providing public goods which increases the demand for this type of goods, hence governments respond by increasing the corresponding supply of open government data.

In addition, our models provide weak support to the hypothesis that open government data is a normal good: that is to say, countries with higher income are associated with higher levels of supply of open government data. However, in our models, political participation, sociodemographic characteristics of citizens, demographic dummy variables and dummy variables capturing the global distribution of income do not help to explain cross-country differences of the supply of open government data.

It is relevant to mention that the main limitation of our analysis is that we use cross section data for our regression analysis which limits the generality of our results. We decided to use data from the GODI index for the year 2016 because this is the most up to date data on GODI. Even if there is data for the Global Open Data Index for other years, the Open Knowledge Foundation has clearly stated that changes in methodology in the calculation of the GODI index make unsuitable the comparison of data between years 2016 and other years. This limits the study of what factors could explain the changes of GODI over time. However, this limitation could be eased as long as more data sets become available in the future that allow other forms of regression analysis such as regression with panel data that might improve the properties of estimation and hypothesis testing as well as the generality of the results.

References

1. Wainwright T, Huber F, Rentocchini F. Open wide? Business opportunities and risks in using open data. eprints.soton.ac.uk; 2014. Available: https://eprints.soton.ac.uk/366901/
2. Attard J, Orlandi F, Scerri S, Auer S. A systematic review of open government data initiatives. Gov Inf Q. 2015;32: 399–418.
3. Hardy K, Maurushat A. Opening up government data for Big Data analysis and public benefit. Computer Law and Security Review. 2017;33: 30–37.
4. Manolea B, Cretu V. The influence of the Open Government Partnership (OGP) on the Open Data discussions. European Public Sector Information Platform. 2013. Available: https://www.europeandataportal.eu/sites/default/files/2013_the_influence_of_the_ogp_on_the_open_data_discussions.pdf
5. Zuiderwijk A, Helbig N, Gil-García JR, Janssen M. Special issue on innovation through open data - A review of the state-of-the-art and an emerging research agenda: Guest editors' introduction. Journal of Theoretical and Applied Electronic Commerce Research. 2014;9: I–XIII.
6. Hossain MA, Dwivedi YK, Rana NP. State-of-the-art in open data research: Insights from existing literature and a research agenda. Journal of Organizational Computing and Electronic Commerce. 2016;26: 14–40.
7. Safarov I, Meijer A, Grimmelikhuijsen S. Utilization of open government data: A systematic literature review of types, conditions, effects and users. Integr Psychiatry. 2017;22: 1–24.
8. Sa C, Grieco J. Open Data for Science, Policy, and the Public Good. Rev Policy Res. 2016;33: 526–543.
9. Zuiderwijk A, Janssen M, Choenni S, Meijer R, Sheikh_Alibaks R. Socio-Technical Impediments of Open Data. ResearchGate. 2012;10: 156–172.
10. Zuiderwijk A, Janssen M. Open data policies, their implementation and impact: A framework for comparison. Gov Inf Q. 2014. Available: https://www.sciencedirect.com/science/article/pii/S0740624X13001202
11. Janssen M, Charalabidis Y, Zuiderwijk A. Benefits, Adoption Barriers and Myths of Open Data and Open Government. Information Systems Management. 2012;29: 258–268.
12. Vetrò A, Canova L, Torchiano M, Minotas CO, Iemma R, Morando F. Open data quality measurement framework: Definition and application to Open Government Data. Gov Inf Q. 2016;33:

325–337.

13. Kučera J, Chlapek D, Nečaský M. Open Government Data Catalogs: Current Approaches and Quality Perspective. In: Kő A, Leitner C, Leitold H, Prosser A, editors. Technology-Enabled Innovation for Democracy, Government and Governance. Springer Berlin Heidelberg; 2013. pp. 152–166.

14. O'Hara K. Data quality, government data and the open data infosphere. 2012 Jul 3. Available: http://eprints.soton.ac.uk/340045/

15. Sadiq S, Indulska M. Open data: Quality over quantity. Int J Inf Manage. 2017;37: 150–154.

16. Faichney J, Stantic B. A Novel Framework to Describe Technical Accessibility of Open Data. ALLDATA 2015: The First International Conference on Big Data, Small Data, Linked Data and Open Data. (IARIA) XPS for publishing; 2015. Available: http://www.iaria.org/conferences2015/ComALLDATA15.html

17. Kapoor K, Weerakkody V, Sivarajah U. Open Data Platforms and Their Usability: Proposing a Framework for Evaluating Citizen Intentions. In: Janssen, M and Mantymaki, M and Hidders, J and Klievink, B and Lamersdorf, W and VanLoenen, B and Zuiderwijk, A, editor. OPEN AND BIG DATA MANAGEMENT AND INNOVATION, I3E 2015. GEWERBESTRASSE 11, CHAM, CH-6330, SWITZERLAND: SPRINGER INT PUBLISHING AG; 2015. pp. 261–271.

18. Böhm C, Freitag M, Heise A, Lehmann C, Mascher A, Naumann F, et al. GovWILD: Integrating Open Government Data for transparency. WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web Companion. Lyon; 2012. pp. 321–324.

19. Peled A. When Transparency and Collaboration Collide: The USA Open Data Program. J Am Soc Inf Sci Technol. 2011;62: 2085–2094.

20. Lourenço RP. An analysis of open government portals: A perspective of transparency for accountability. Gov Inf Q. 2015;32: 323–332.

21. Gurin J. Open Governments, Open Data: A New Lever for Transparency, Citizen Engagement, and Economic Growth. SAIS Review of International Affairs. 2014;34: 71–82.

22. Ahmadi Zeleti F, Ojo A, Curry E. Exploring the economic value of open government data. Gov Inf Q. 2016;33: 535–551.

23. Alt R, Franczyk B. Business Models in the Data Economy: A Case Study from the Business Partner Data Domain. In: Alt R, Franczyk B, editors. Proceedings of the 11th International Conference on Wirtschaftsinformatik (WI2013). Leipzig: Universität Leipzig; 2013. p. 15.

24. Hansen HS, Hvingel L, Schrøder L. Open Government Data – A Key Element in the Digital Society. Technology-Enabled Innovation for Democracy, Government and Governance. Springer Berlin Heidelberg; 2013. pp. 167–180.

25. Susha I, Grönlund A, Janssen M. Driving factors of service innovation using open government data: An exploratory study of entrepreneurs in two countries. Information Polity. 2015;20: 19–34.

26. Zuiderwijk A, Helbig N, Gil-Garcia JR, Janssen M. Special Issue on Innovation through Open Data-A Review of the State-of-the-Art and an Emerging Research Agenda. Guest Editors' Introduction Journal of theoretical and applied electronic commerce research. 2014;9: 1–2.

27. Lin Y. Open data and co-production of public value of BBC Backstage. INTERNATIONAL JOURNAL OF DIGITAL TELEVISION. 2015;6: 145–162.

28. Juell-Skielse G, Hjalmarsson A, Johannesson P, Rudmark D. Is the Public Motivated to Engage in Open Data Innovation? Lecture Notes in Computer Science. 2014. pp. 277–288. doi:10.1007/978-3-662-44426-9_23

29. Chan CML. From Open Data to Open Innovation Strategies: Creating E-Services Using Open Government Data. IEEE; 01/2013. pp. 1890–1899.

30. Eckartz S, Broek T van D, Ooms M. Open Data Innovation Capabilities: Towards a Framework of

How to Innovate with Open Data. In: Scholl HJ, Glassey O, Janssen M, Klievink B, Lindgren I, Parycek P, et al., editors. Electronic Government. Springer International Publishing; 2016. pp. 47–60.

31. Worthy B. THE IMPACT OF OPEN DATA IN THE UK: COMPLEX, UNPREDICTABLE, AND POLITICAL. Public Adm. 2015;93: 788–805.

32. Maier-Rabler U, Huber S. “Open”: the changing relation between citizens, public administration, and political authority. 1. 2011;3: 182–191.

33. Bates J. The strategic importance of information policy for the contemporary neoliberal state: The case of Open Government Data in the United Kingdom. Gov Inf Q. 2014;31: 388–395.

34. Ruijer E, Détienne F, Baker M, Groff J, Meijer AJ. The Politics of Open Government Data: Understanding Organizational Responses to Pressure for More Transparency. Am Rev Public Admin. 2020;50: 260–274.

35. Ubaldi B. Open Government Data Towards Empirical Analysis of Open Government Data Initiatives. 2013. doi:10.1787/5k46bj4f03s7-en

36. Jetzek T, Avital M, Bjørn-Andersen N. Generating Value from Open Government Data. ICIS 2013 Proceedings. aisel.aisnet.org; 2013. Available: https://aisel.aisnet.org/icis2013/proceedings/GeneralISTopics/5/

37. Jetzek T, Avital M, Bjørn-Andersen N. Generating value from open government data. International Conference on Information Systems (ICIS 2013): Reshaping Society Through Information Systems Design. Milan; 2013. pp. 1737–1756.

38. Attard J, Orlandi F, Auer S. Value Creation on Open Government Data. 2016 49th Hawaii International Conference on System Sciences (HICSS). 2016. doi:10.1109/hicss.2016.326

39. Oakland WH. Chapter 9 Theory of public goods. Elsevier; 1987. pp. 485–535.

40. Cornes R, Sandler T. The Theory of Externalities, Public Goods, and Club Goods. 1996. doi:10.1017/cbo9781139174312

41. Hettich W, Winer SL. Democratic Choice and Taxation. 1999. doi:10.1017/cbo9780511572197

42. Mueller DC. Public Choice III. 2003. doi:10.1017/cbo9780511813771

43. Hankla C, Martinez-Vazquez J, Rodríguez RP. Local Accountability and National Coordination in Fiscal Federalism. 2019. doi:10.4337/9781788972178

44. Roemer, J.E. (2001). Political Competition: Theory and Applications. Cambridge: Harvard University Press.

45. Bates J. The Domestication of Open Government Data Advocacy in the United Kingdom: A Neo-Gramscian Analysis. Policy & Internet. 2013. pp. 118–137. doi:10.1002/poi3.25

46. Dos Santos Brito K, Da Silva Costa MA, Garcia VC, De Lemos Meira SR. Assessing the benefits of open government data: The Case of meu congresso nacional in Brazilian Elections 2014. In: Zhang J. KY, editor. ACM International Conference Proceeding Series. Association for Computing Machinery; 2015. pp. 89–96.

47. Purwanto A, Zuiderwijk A, Janssen M. Citizen engagement with open government data. Transforming Government: People, Process and Policy. 2020. pp. 1–30. doi:10.1108/tg-06-2019-0051

48. Hong S. Electoral Competition, Transparency, and Open Government Data. The 21st Annual International Conference on Digital Government Research. 2020. doi:10.1145/3396956.3398254

49. Kochi I, Rodríguez RAP. VOTING IN FEDERAL ELECTIONS FOR LOCAL PUBLIC GOODS IN A FISCALLY CENTRALIZED ECONOMY. Estud Econ. 2011;26: 123–149.

50. Bergstrom TC, Goodman RP. Private Demands for Public Goods. Am Econ Rev. 1973;63: 280–296.

51. Beron KJ, Murdoch JC, Vijverberg WPM. Why Cooperate? Public Goods, Economic Power, and the Montreal Protocol. Review of Economics and Statistics. 2003. pp. 286–297. doi:10.1162/003465303765299819

52. Jackson PM, Atkinson AB, Stiglitz JE. Lectures on Public Economics. The Economic Journal. 1981. p. 573. doi:10.2307/2232622

53. Tresch RW. Public Finance: A Normative Theory. Academic Press; 2002.

54. Mas-Colell A, of Economics Andreu Mas-Colell, Whinston MD, Green JR, of Political Economics Jerry R Green. Microeconomic Theory. Oxford University Press, USA; 1995.

55. Rubinfeld DL. The economics of the local public sector. Handbook of public economics. Elsevier; 1987. pp. 571–645.

56. Mason H, Wiggins C. A taxonomy of data science. Dataists com. 2010.

57. Kotsiantis SB, Kanellopoulos D, Pintelas PE. Data preprocessing for supervised leaning. International Journal of Computer Science. 2006;1: 111–117.

58. Liu H, Motoda H. Feature Extraction, Construction and Selection: A Data Mining Perspective. Norwell, MA, USA: Kluwer Academic Publishers; 1998.

59. Garg A, Tai K. Comparison of statistical and machine learning methods in modelling of data with multicollinearity. Int J Model Ident Control. 2013;18: 295–312.

60. George G, Osinga EC, Lavie D, Scott BA. Big Data and Data Science Methods for Management Research. AMJ. 2016;59: 1493–1507.

61. John Lu ZQ. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. J R Stat Soc Ser A Stat Soc. 2010;173: 693–694.

62. Cao L. Data Science: Challenges and Directions. Commun ACM. 2017;60: 59–68.

63. Salimans T. Variable selection and functional form uncertainty in cross-country growth regressions. J Econom. 2012;171: 267–280.

64. Greene WH. Econometric analysis. Pearson Education India; 2003.

65. Long JS, Ervin LH. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. Am Stat. 2000;54: 217–224.

66. Agunbiade DA, Adeboye NO. Estimation of Heteroscedasticity Effects in a Classical Linear Regression Model of a Cross-Sectional Data. Progress in Applied Mathematics. 2012;4: 18–28.

67. Zhu L, Fujikoshi Y, Naito K. Heteroscedasticity checks for regression models. Sci China Ser A Math. 2001;44: 1236–1252.

68. Rao CR, Toutenburg H. Linear Models. Springer Series in Statistics. 1995. pp. 3–18. doi:10.1007/978-1-4899-0024-1_2

Productos generados

El resultado de este reporte tecnico se sometera a dictaminacion para su futura publicacion en una revista internacional indexada y con factor de impacto aceptada por el CONACYT.