3rd International Conference

on Public Policy (ICPP3)

June 28-30, 2017 – Singapore

Panel TO2P35 Session 1

The Politics of Open Government in Iberoamerican Countries: A Comparative View About its Determinants and its Implementation Processes

Title of the paper

Assessment of the open government data impact in Latin American Countries: Key factors exploratory analysis

Author(s)

Abella, Alberto, Desidedatum, Spain, alberto.abella@meloda.org

Date of presentation

Wednesday, June 28th 16:15 to 18:15 (Block B 3 - 3)

Abstract:
The idea of open government cannot be understood without considering the systematic publication of data about public administrations. Thus public open data portals become critical resources to ensure transparency and accountability. The interaction of citizens and organizations with these portals depends on organizational, economic, technical and social factors that must be analysed to assess the use of these resources and their efficient use.

The paper presents results of sampling on these portals and the ecosystems of organizations around them. These relevant findings may help in redefining strategies for releasing public data, coordinating them and to prioritize investments in these resources.

Keywords: *Open government, Open data, Apps, MELODA, information reuse*

1 Introduction

The massive popularisation of digital technologies during the last decade and the increasingly availability of internet connections have extended the ability of governments to interact directly with the citizens or groups of them, even beyond the rigid and less frequent participation tools implemented in the current legislation.

Open government is increasingly demanded by a new groups of citizens (Ramirez-Alujas & Dassen, 2014) taking advantage of some of these changes. Their demands include a more active participation, a dramatic increase of the accountability of the public policies, and urge public managers to take into account citizen's opinions in the design and implementation of new policies.

This article reviews the role played by the government publication of data in the context of open government policies. The motivational factors for data publication are identified and a model for the assessment of the data reuse impact is presented. The model identifies different categories of data reusers and the role played by the infrastructures as enablers of the data publication.

Next section presents a methodology for the exploratory analysis. This methodology covers the analysis of infrastructure, data and services. The analysis of infrastructure describes a simplified model for the assessment of the maturity of the different open data portals. The methodology for the exploratory analysis of data covers the analysis of their reusability factors according to the metric MELODA. (Abella et al. 2014) Finally, the analysis of apps and services identities the elements explored and evaluated.

The third section of the article presents the results and interpretation of the exploratory analysis for the three elements, infrastructures, data published and data-driven services. Last section, conclusions, describes the main findings and their potential relations. It also points out some recommendations for open data managers in the area and how to move ahead with the analysis.

2 Data context of open government

The Open Government Partnership[1] (OGP) is possibly the most relevant Open Government organisation in the world. In 2017, 14 out of the main 19 Latin American countries analysed belong to such organisation as identified in illustration 1 and table 1.

The final goal of open government is to provide a more efficient and fair governance for the citizens. Currently, whenever we talk about open government we refer to three main points (Ramirez-Alujas, 2014):

1) Improving accountability, transparency and access to information (FOI)

2) Extensive ability to reuse Public sector information and

3) Facilitating the participation of citizens in the design and Implementation of public policies through different mechanisms

The impact of these three elements are deeply interlinked making it hardly quantifiable individually.

2.1 Transparency moves to data

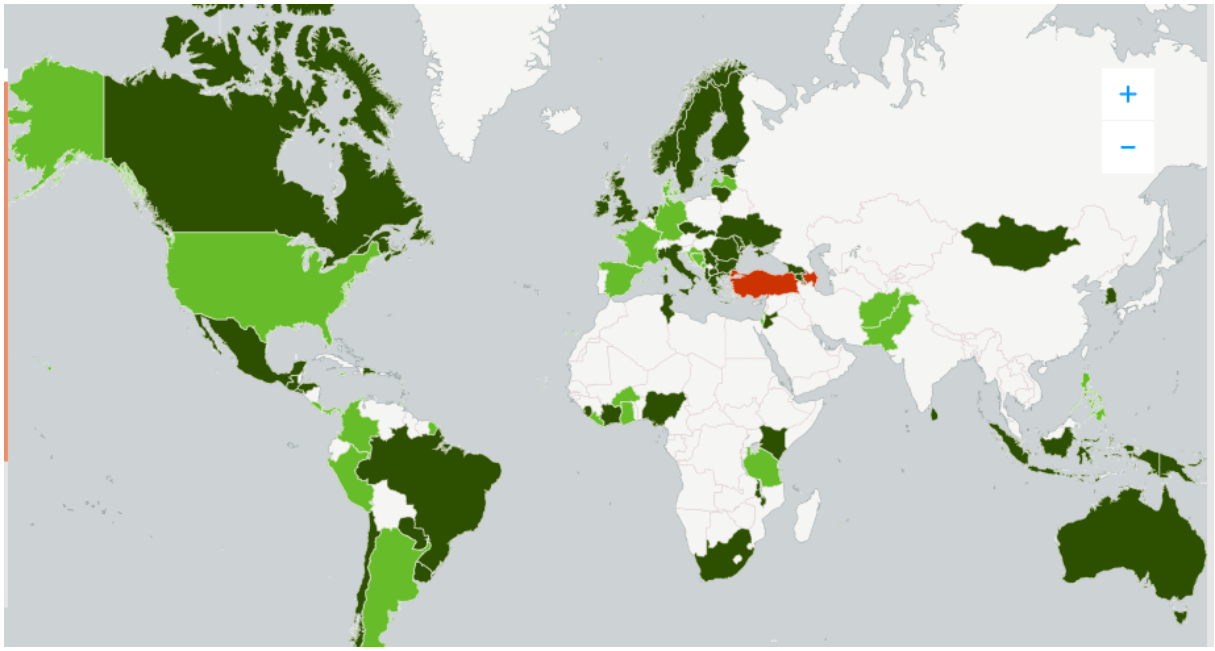
Transparency standards are moving from snapshot to emerging league-based ones. Snapshot standards means that the standard defines well documented criteria to be implemented and assessed, together with a periodic review and assessment (sometimes longer than 2 years) of the implemented criteria run by the supporting organisation of the standard. These standards' assessment procedures have the advantage of a rigorous and accurate assessment of the evidences supporting the transparency standard, but on the other hand it leaves most of the time without any check about the accomplishment of the required transparency criteria and requires from a mostly manual assessment. Conversely league-based transparency standards trust on a continuous assessment of the transparency through published data. In order to provide a continuous assessment automation tools have to be used in order to manage the huge amount of data. On the other hand this automation set some limits on the assessment.

Accordingly to this trend transparency requirements are moving from static documents to streams of data showing the compliance with the requirements. Thus the infrastructures to comply with these transparency requirements are moving from just a collection of documents listed in the

1http://opengovprtnership.org

suitable sections of a web page, to a multiple source of data embedded in portals providing additional clarification for the data released and inviting the users to reuse and to process them.

Illustration 1: Worldwide Open Government Partnership members 2017



Source: Open Government Partnership site. (Retrieved May 2017)

2.2 Data publication underlies open government

The first element of open government, Improving accountability, transparency and access to information (Ramirez-Alujas, 2014), impacts on the final goal of open government by creating barriers to unfair policies, limiting corruption and asking for responsibilities based on the results of the implemented policies.

Besides this, the public administration is urged to publish data in order to demonstrate its probity and transparency. Thus, data are critical for the design of new services in order to meet the expectations of the different groups of citizens (Abella, 2017). Finally, It is needed good data collection mechanisms in order to analyse public opinion.

Independently of the pillar of open government involved, there is a strict necessity of accountability of public policies and resources. Therefore whenever an open government policy is implemented, an underlying layer of data management has to be created.

That intense publication of data depends on a proper internal data management and governance. Needless to say, an intense open government approach requires that data governance, the technique to make data available when and where it is required, turns increasingly

important. So, in a close future, those pubic entities embracing open government principles will have to assign the role of the CDO (Chief Data Officer), responsible person for data governance.

Table 1: Latin American members of OGP

Country	OGP membership	Country	OGP membership
Argentina	yes	Honduras	yes
Bolivia	no	Mexico	yes
Brazil	yes	Nicaragua	no
Chile	yes	Panama	yes
Colombia	yes	Paraguay	yes
Costa Rica	yes	Peru	yes
Cuba	no	Dominican Republic	yes
Ecuador	no	Uruguay	yes
El Salvador	yes	Venezuela	no
Guatemala	yes		

Source: OGP members list. (Retrieved May-2017)

However this internal perspective of data management could distract for the effects of data publication on the surrounding society. This work pays attention to the use of these data and the derivate impacts produced whenever these data are reused thank to their release with an open licence. What is commonly known as open government data.

2.3 The ecosystem of data reusers

Open government data are those data released by public administrations with a licence allowing to reuse the data by anyone and with any purpose, with the only requirement of attributing the source of the data. (Open Knowledge International, 2015) . These data, are usually released through *open data portals*. These portals are the connection between the internal data management systems of a public administration and the reusers of the data.

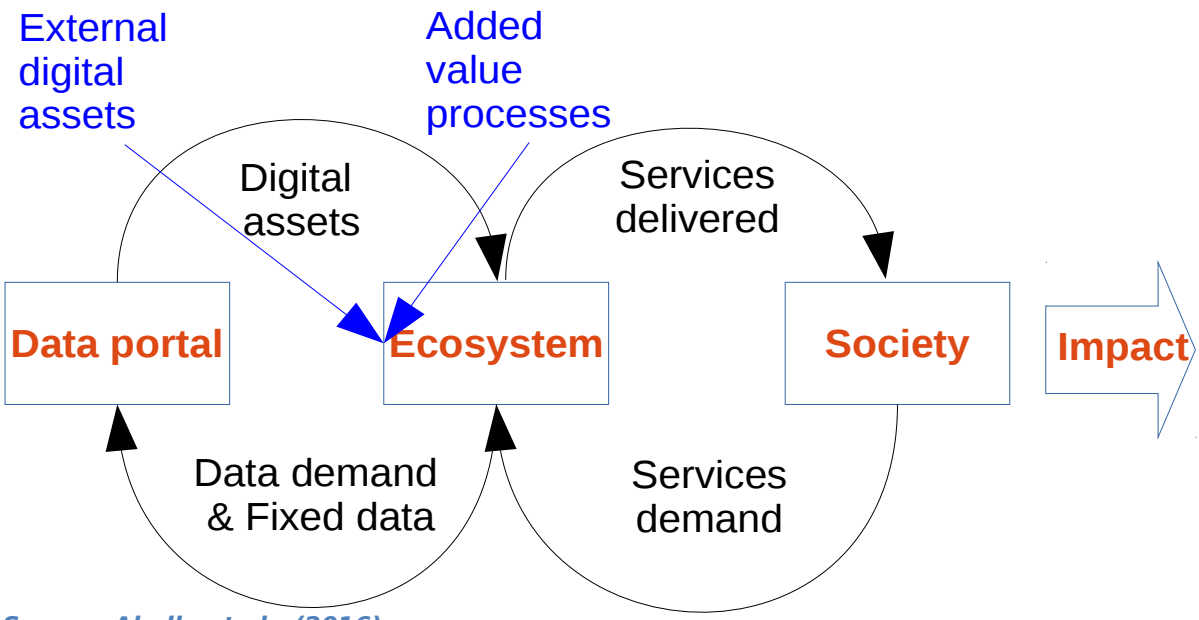
Surrounding those portals the data released could nurture a lively ecosystem of reusers. They exploit the value of the data released by mixing them with other data or by using added value processes as described in illustration 2.

In illustration 2 the elements of this ecosystem are identified. On the left is the open data portal releasing constantly data to the ecosystem, and also collecting feedback from the reusers, new requirements for existing data, new datasets to be released and also fixing some of its data thanks to the feedback collected from reusers.

The data portal publishes data, but also some standardisation documents, detailed mechanisms to access the data, code examples to access the data, etc. All these resources are grouped under the term digital assets in the illustration 2.

The ecosystem in the middle of the illustration comprises the agents reusing the data. They range from other public institutions to private entities for profit or non-for-profit and sometimes active citizens (developers, data journalists, etc.). Further details about reusers could be found in section 2.5 and illustration 3. These reusers deliver services to the society. Society is the next element of the illustration. It represents not only citizens but also companies, organisations, and even the public administrations as long as they were clients of the services and products created by the ecosystem. The society claims for new services or improvements of the existing ones to the ecosystem, which, eventually, boost the demand for new or better data releases.

Illustration 2: Data ecosystem around a data portal.



Source: Abella et al., (2016)

The last element of this illustration is the impact on the society caused by the released data. Jetzek et al. (2013) identified four impact mechanisms. First one, innovation. Products and services are created or enriched with the data released and it allows the creation of innovative products and

services. The second mechanism is transparency. Transparency accounts for the increasing competition, limited corruption and new opportunities due to a better knowledge of the requirements and demands of the public administrations. Third impact mechanism is the efficiency. The data released make the same administration (and others) more efficient by optimising processes and eliminating redundant ones. Last mechanism is participation. The data publication increases the ability of the citizens to join initiatives and to participate (Linders and Wilson 2011) and increases citizen engagement (Kogan, 2014).

2.4 Data management in public administrations

The open data portal is the icing of the cake in the process of publishing the data in a public organisation. In order to succeed in data publication and reuse, a public organisation (and mostly a private one as well) has not only to have a portal, but also to perform these tasks (non exhaustive list):

- Prioritise the different available data to be released, balancing the needs for transparency, the interest for the reusers of the data released and the resources involved to release these data.
- Define responsibilities for the data collection and approval for publication
- Define a legal framework that allows the sustainability of the publication data initiative and demonstrate the support of the public entity managers.
- Implement technical processes that extracts the right data from the internal information systems and transform them into some datasets suitable to be reused.
- Implement technical procedures that automates, as much as possible, the publication processes.
- Define communication channels for data reusers in order to collect, analyse and deal with the feedback.
- Define an analysis resource in order to assess the actual reuse and to compare it with the expectations and to identify the impact on society.

2.5 The reusers of data

Illustration 3 describes the different categories of data users and reusers. On the left is identified the publishing entity, which at the same time used to be one of the main consumers of the data released (category 5 in illustration 3). In fact, it is frequent the use of open data portals, as internal breaking-silos tool. Silos are meant, the information systems managed individually and unconnected inside a public organisation.

It is also true that sometimes connected public entities (i.e. municipalities of a regional government) are important reusers of the data released (category 6).

The direct reusers are those entities that take the data in order to create something different and that they are not their final users. Those reusers are classified into two categories, those whose goal is social and non-for-profit(category 3), and those looking for a profit (category 2).Among this last category the most popular types of companies are, according to (Red.es, 2017) :

- Market research and advertising
- Research and consultancy
- Publishers (books and other databases)
- Geolocated information publishers
- Economic and financial risk analysis companies
- Cultural companies

They present a huge variety of business models, freemium, advertising, etc. that has been analysed in (Zeleti et al, 2016).

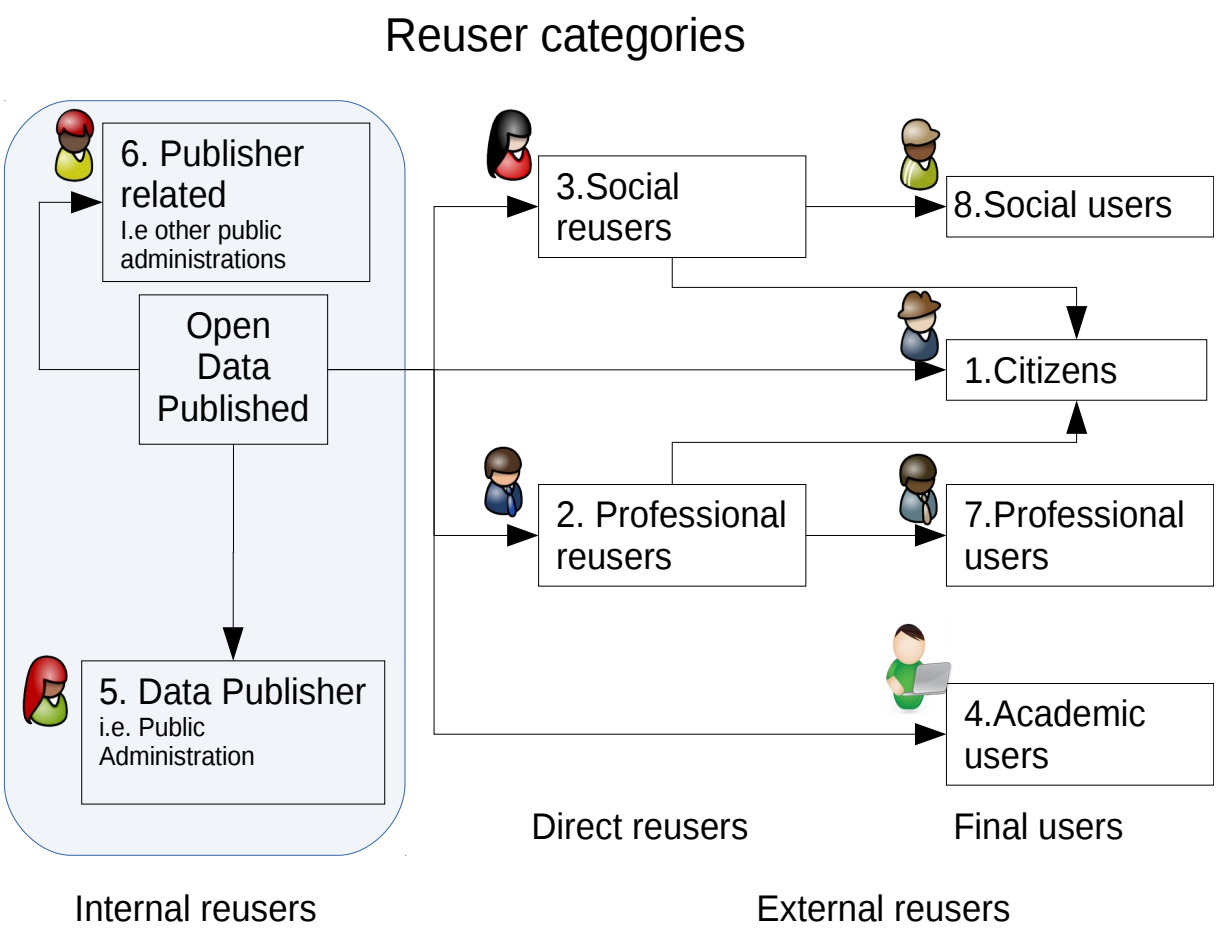
However it is also to be mentioned that these categories are not strict divisions and that some companies could provide services in several of them. Additionally, according to Red.es (2017), it is common that these companies also perform some other business activities different to those related with data reuse.

On the right hand part of the illustration are identified the final users. They are divided into four categories. First one are the regular citizens (category 1), they sometimes access directly the published datasets or sometimes they use some other services provided by for-profit or non-for profit suppliers. This category includes two very special user types that are the data journalists who use the data for creating their contents and also the

individual developers, those innovators acting individually, previously to launch their businesses or social initiatives.

Next category are direct users of the social services provided by category 3. A common example of this category 8, are associations of disabled people that use accessibility points from public data sources compiled by other more technical social entities (category 3). They provide consume that data-driven service as an added value for their members.

Illustration 3: Reusers' categories



Source: Abella (2017)

Category 7 accounts for those professional users using the services created by for profit entities (category 2). They are the regular customers of the companies listed in previous paragraph.

Last but not least are the academic reusers (category 4) which for their special use requirements (i.e. semantic requirements, standardisation, volume, etc.), they have to be counted separately.

All these users of data have their own requirements for the data released. So, the open data portals will have to provide solutions in order to meet the initial goal beyond the mere data publication which is to generate an impact on the surrounding society, independently of the impact is measured in terms of transparency, new services created base on the released data, increase efficiency of the own open data publishers, etc. Further modelling of this impact can be found in Abella et al., 2017.

2.6 The role of the open data portals

According to the illustration 2, data are released from the original source through the open data portals in order to feed regularly an ecosystem of reusers. However, in order to be effective allowing the reuse of data, what are the features that a open data portal should implement?. These features are what is identified in the market as Data Management Solutions (DMS). Here there is a list of them most notable ones.

Need 1: Data publication. In order to meet this requirement the DMS has to publish the data as soon as it is approved by the administrator in the original source and make it available properly for the reusers with a group of additional information namely metadata (owner, update frequency, last update, topic, tags etc). Every dataset published has to fill them and to be able to customise a group of metadata describing the data released.

Additionally, other technical formats, different from the original one could be useful for some reusers. Therefore, It is an interesting feature that the DMS would be able to transform the original dataset into different formats. Initially, the more the better, however the experience shows that csv and json are the most widely used for regular data.

Need 2. Data search. Once published, the amount of datasets, thousands eventually, could make difficult to find out the interesting data for every reuser. Thus, some search facilities has to be available for the users, not only for the final users but also for the automatic search through some automatic mechanism. This feature should include also some search engine optimisation (SEO) features for making the published data popular for search engines.

Need 3: Data harvesting. Data have to be updated from the original source frequently, ideally whenever a change occurs. The DMS has to be able to create processes that make this task a seamless integration into the original information systems. Otherwise manual update of the data would impose a painful task demanding additional resources from the organisation.

Need 4: Permissions and workflow. Some of the fields belonging to a dataset could have some privacy or security limitations. Consequently, some restrictions to access to them have to be put in place. Removal of these fields could be enough for external uses, but if the DMS also provides service inside the organisation, an advanced permission systems is required. The more granular the better.

Need 5: Users synchronisation. The owner of the published data has to have the rights to correct, upload, unpublish, comment, etc., their data. So, whenever the organisation is complex and lively, if the users are not synchronised with a corporate identity tool, user permissions will become obsolete quickly and they will turn into a restriction for data publication.

Need 6: Visualization. Although raw data is right for industrial users, regular citizens (category 1 in point 2.5), require from some visualisation help in the form of graphs, maps, diagrams, etc. It will help to make the data understandable and demanded for these categories of reusers.

Need 7: Processing. Although ideally the data will be grabbed just in the format required to be published, this case is quite uncommon. Some kind of filtering, removal of empty records, transformation of the original data would be needed. Accordingly, if the DMS has some tool for treating the data it will help to increase the quality of what is published and to accelerate their publication.

Need 8. API and automated access to data. An Application Program Interface will allow the automated consumption of the published data and it will reduce useless downloads for full datasets. This API has to be available as soon as the data is uploaded and approved for publication. API has to be properly documented and with some snippets of code. API could be customised for every user and it should have some kind of flow limit.

Need 9. Feedback channel. Publication is one step in the lifecycle of data, and therefore a feedback channel has to be available, not only from the technical tool but also from a supportive organisation. DMS has to provide simplicity to manage such feedback and to assign feedback messages to the right people and to manage that all of them are properly treated.

Need 10. Analytics of the access. DMS has to provide tools for the analysis of the datasets consumption, publication, reuse, etc. It has to identify, voluntary, to those reusers willing to be identified, also to analyse the collected feedback and the access mechanisms to the data.

Table 2: Importance of the requirement of a DMS for the different data reusers' categories

Requirement	Category of data reusers (according to illustration 3)							
	1	**2**	**3**	**4**	**5**	**6**	**7***	**8***
Data publication	xx			x	x	xx		
Data search	x	xx	xx	xx		x		
Data harvesting		xx	xx		xx			
Permissions and workflow					xx			
Users synchronisation		x	x		xx			
Visualization	xx	x	x	x		x		
Processing		xx	xx	x	x	x		
API	x	xx	xx	xx	x	x		
Feedback	x	xx	xx	xx		xx		
Analytics				x	xx	x		
Real-time		xx	x	x	xx	xx		

Legend: xx very important. x important.

*** Those categories do not interact directly with the open data portal**

Source: Own elaboration

Need 11. Real-time access to data. Although data harvesting has been already identify as a need, there is an increasing need for the connection in real-time with the sources. This need imposes qualitative changes into the design of the tools and redefines concepts like version, dataset, etc. Gradually, static datasets with versions turn into streams of data, also named as datajets (Abella, 2017).

2.7 The open data portals in Latin America

Open data portals are becoming increasingly popular in Latin America as shown in illustration 4 where more than 250 are identified. For the shake of this analysis only 212 will be taken into account for being located in the main Latin American countries according to the table 3. Note that having lots of portals could be a symptom of a decentralised administrative organisation more than an extensive implementation of open data policies.

Table 3: Open data portals listed for main Latin American countries

Country	# OD portals	Country	# OD portals
Argentina	20	Honduras	1
Bolivia	30	Mexico	16
Brazil	22	Nicaragua	2
Chile	14	Panama	4
Colombia	36	Paraguay	4
Costa Rica	13	Peru	19
Cuba	3	Dominican Republic	3
Ecuador	4	Uruguay	7
El Salvador	6	Venezuela	3
Guatemala	4		

Source: Datainception.io. Retrieved May-2017

First of all the accountability of the public institutions requires to publish information freely available for the citizens. This information was in the past mostly published as static information (documents). It is increasingly changing into more real time and interactive data sources where citizens and other organisations are able to retrieve the requested data.

Another motivation for the release of open government data is to increase the efficiency of public administrations. Sharing data increase the performance of public administration’s processes and enables the creation of innovative and more efficient services for the citizens and the surrounding society (Jetzek et al., 2014).

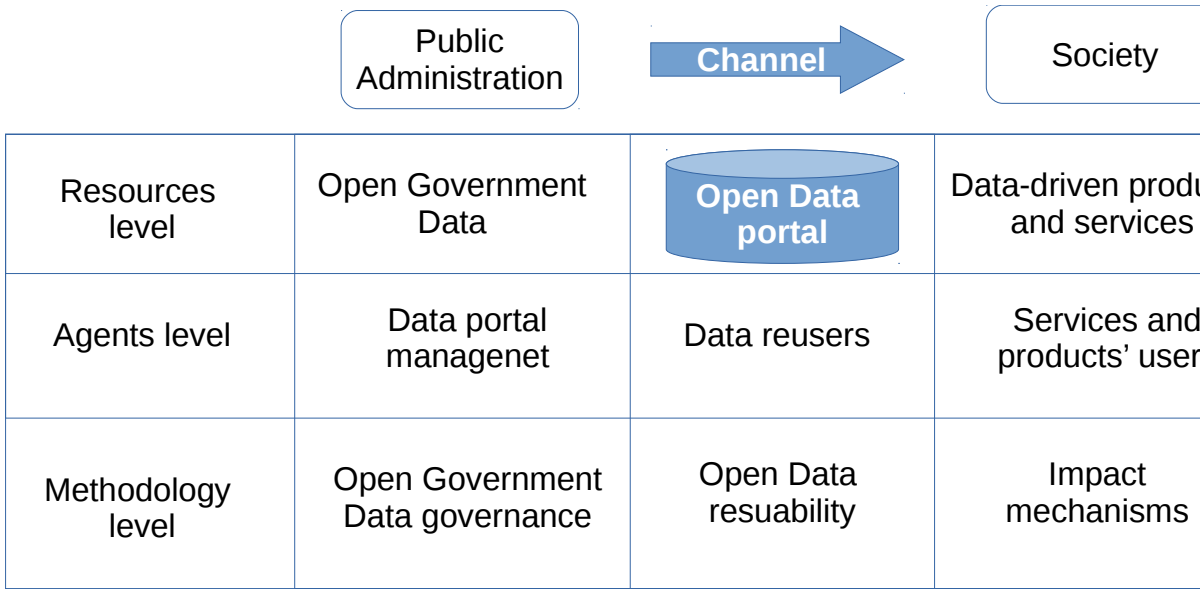
Illustration 4: Open Data portals in Latin American countries

Source: Datainception.io. Retrieved May-2017

3 Methodology for the analysis

Illustration 5 depicts the master elements of the analysis carried out in this article. Open government, as stated in section 1, looks for the impact on the surrounding society. There are three levels. The first level identifies actual objects the data publish through the open data portal that turns into data-driven products and services consumed by the society. The second level is the agents level. Agents mean the people managing and using the data. Internally on the public administration are the responsible people for the collection and data publication. The data reusers are the people in the channel creating the products and services that are finally consumed by the last group of agents, the final users. The third level identified the methodologies required for the management of data wherever it is found. Inside the public organisations it should be implemented data governance principles (see section 2.4). Whenever it is published it should be managed according the principles of reusability (see section 3.2.1) and once in the society it should assessed according to some model (Jetzek et al., 2013) (Abella et al., 2017).

Illustration 5: Master elements for the impact of open government data



Source: Own elaboration

An extensive and statistically sound sampling of the full open data ecosystem in Latin America exceeds the scope of this study. Instead of this, it has been performed an exploratory analysis on a limited universe in order to detect some of the most remarkable aspects before quantifying them.

For this purpose, the central open data ecosystems surrounding the open data portal in every country of Latin America has been chosen as universe for the analysis. It comprises open data portals, their published datasets and the applications and services created based on these data. Thus the global amount of analysed countries were initially nineteen as listed in table 1 in section 2.2.

3.1 Methodology for exploratory analysis of open data portals

From the original list of nineteen countries, two of them do not have something similar to an open data portal (Nicaragua and El Salvador) and 2 of them were not actual open data portals but central statistics sites (Cuba and Venezuela). In the last case data were available only under payment. So, only 15 open data portals, were analysed.

For each sampled portal the following values were identified:

- Use of a specific Data Management System tool.
- Number of published datasets.
- Availability of mechanisms for publishing data updates.
- Availability of a catalogue of data resources,
- Existence of mechanisms of automated connection with the data (API).
- Availability of a section in the portal that identifies data-driven services, based on the data of the portals
- Number of applications and services listed in the portal section

3.1.1 Simplified Maturity model for open data portals

In order to explore the maturity of the open data portals a simplified maturity metric has been developed based on the one proposed by Carrara et al. (2016). The categorization of the results follows the same group of categories as the MELODA data reusability metric (Abella et al., 2014).

The analysed aspects and their importance are presented in table 4, introducing the following elements into consideration:

- Whether the population of datasets exceeds 30 elements.
- The availability of a feed of the updates of the data.

- The availability of an API that allows automated access to data by external users.
- The use of a specific DMS tool for the publication of data. CKAN, Socrata publica, DKAN, OpenDataSoft, Junar and ESRI Open Data tools have been considered as DMS for this work.
- The Availability of a section in the portal that identifies data-driven services.

With the information obtained, an analysis of the maturity of the portal data has been carried out according to the weights listed in table 4.

Table 4: Elements of the simplified model for Open data portal maturity

Concept	**Concept of Pan-European report**	**Weight**
Having more than 30 *datasets*	*Spread of data*	20%
Having a *feed* with updates of the catalogue	*Usability of the portal*	10%
Use of DMS	*Usability of the portal*	15%
API availability	*Re-usability of data*	25%
Apps/services portal	*Re-usability of data*	30%

The first concept having more than 30 datasets reflects the maturity on finding relevant data to be published and it is a threshold that make manual update of information hardly bearable in the mid-term.

The second concept, having a feed, or other mechanism that allows not to be recurrently retrieving full catalogue in order to detect when a dataset is updated. Otherwise reusers should be requesting constantly the catalogue of data in order to detect changes.

The third point includes the use of a specific tool (DMS) for publishing the data. Even having such tools publishing regularly data and feeding a lively ecosystem of reusers is a intensive task. Without it is simply unsustainable in the long term.

The forth element to be considered is the availability and use of an API or other automation mechanism (i.e. SPARQL endpoint) to make available the

data in an industrial way. Full download of datasets makes extensive use of resources (transmission, storage, etc.) when need for the full content of datasets is a rare event. Sometimes DMS solutions have the capability of automated access but this automated capability is not made public for reusers.

Last element denotes having a space or section where agents reusing the published data can show their data-driven services and apps, demonstrating the impact on the society.

3.2 Methodology for exploratory analysis of datasets

The concept of dataset is studied by Renear, Sacchi and Wicket (2010) but by that time open data portals were not popular if actually exist. In order to provide context to this statement by the end of 2010, it was launched the first open data portal in Spain (Open Data Euskadi[2]), conversely nowadays there there are more than 150 listed[3].

Reuse is defined as “To use something again.” Reusing information has been studied in many disciplines, including financial sector (Chan; Greenbaum; Thakor, 1986), clinical medicine (Kim, 2005), and scientific information (Faniel & Jacobsen, 2010).

This work defines reusable dataset those that meet four minimum conditions that facilitate its reuse:

1) Absence of technical barriers to reuse;

2) Possibility of automated access to information (Eaves, 2010);

3) Existence of a legal framework that allows its use (Eaves, 2010);

4) Access to knowledge of the structure of the information published.

Taking into account the breadth of the universe under study, more than 20.000 datasets (see table 7), the exploratory analysis just sampled randomly 30 datasets out of them. It provides a confidence interval of 18 points at 95%. Thus definitive statements cannot be made and results can only show some trends and qualitative results.

For each dataset, its topic has been identified according to the DCAT-AP standard (European Union, 2015) as well as each of the six dimensions of the MELODA metric (Abella et al, 2014, 2017) as described in section 3.2.1. Other fields collected were the dataset name and the access URL.

2http://opendata.euskadi.eus

3http://mapa.datos.gob.es

3.2.1 Data Reusability : key to impact

In order to quantify these conditions, MELODA[4], a metric for data reusability has been defined (Abella et al., 2014, 2017) and it will be summarised here in order to provide a simple view of the provided results in section 4.

The datasets are analysed according to six dimensions: the legal licensing to use the data, the technical standard use to stored the data, the access mechanism a user should implement to reach the data, the data model (the meaning of the different fields of the dataset), its geographic content and its update frequency.

For every one of these six dimensions there are defined different levels of achievement, and a weight associated to every level. Composing these weights according to the formula in the lower part of illustration 6, a final figure is obtained. Values of such figure are grouped into four categories in order to provide simpler analysis. The groups can be found in table 5.

Example of the application of MELODA 4.13 metric

Junta de Castilla y León, a regional government in Spain, publishes a dataset with the Youth Accommodation resources[5], as a XML file but with its own data model that they make public, licensed just for attribution (creative commons 3.0) and with a direct link to download the full dataset. The dataset includes coordinates and some text fields in order to geolocate the resources and it is updated once a day.

The corresponding levels in each dimension are:

Legal: Level 5 (Attribution)	100%
Technical standard: Level 4 (open standard with metadata)	100%
Access: Level 3 (unique URL parameters to access dataset)	50%
Data model: Level 3 (Own ad hoc data model published)	50%
Geolocation information : Level 5 (Full geographic information)	100%
Data model: Level 3 (Hours)	70%

4 Further information and examples can be found at http://meloda.org and the full description of the metric in the link http://www.meloda.org/full-description-of-meloda/

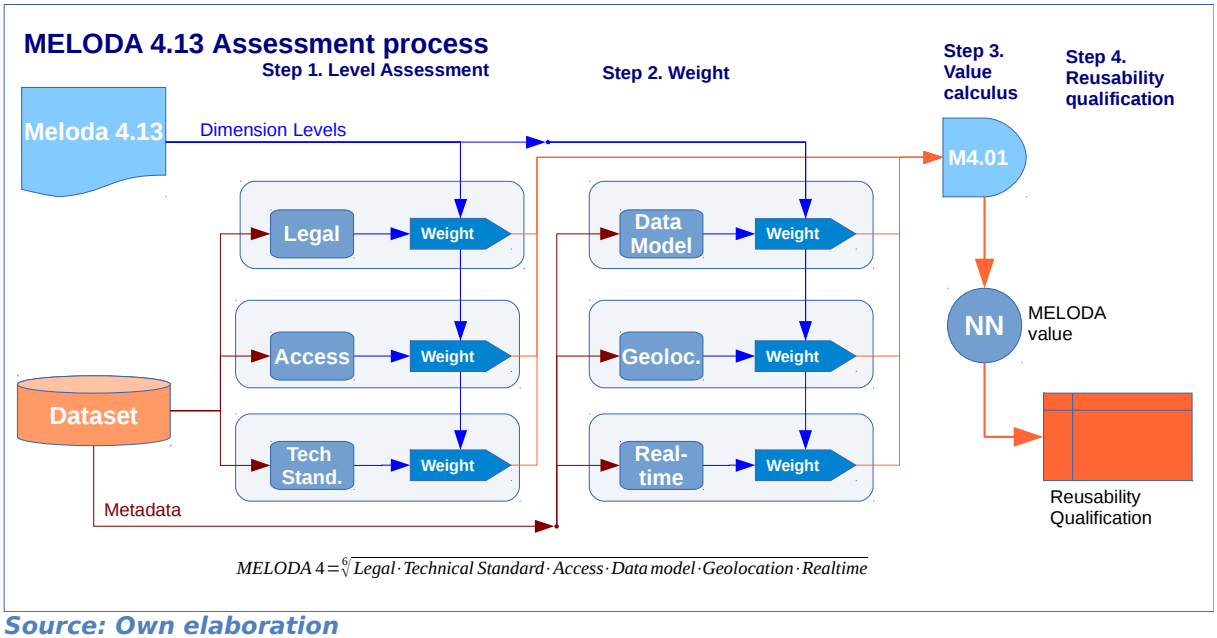
5http://www.datosabiertos.jcyl.es/web/jcyl/set/es/sociedad-bienestar/AlberguesJuveniles/1284198756067

Result

Meloda = $\sqrt[6]{100 \cdot 100 \cdot 50 \cdot 50 \cdot 100 \cdot 70}$ =74.8% Reusability : Good

Symbol:

Illustration 6: MELODA 4.13 assessment process and calculation formula



Source: Own elaboration

Table 5: Qualification table for MELODA values

MELODA 4 range	100-75	75-50	50-25	25-0
Reuse qualification	Optimum	Good	Basic	Deficient
Symbol				

Source: Own elaboration. Available at http://meloda.org

3.3 Methodology for exploratory analysis of apps and services

This exploratory analysis has identified apps and services based on the published datasets only for six portals (Brazil, Mexico, Panama, Paraguay, Peru and Uruguay). out of the fifteen with data available. The universe of service comprises 209 apps and services identified on these portals. A random sample of 20 services where chosen.

These are the data extracted for each one of these services:

- The topic of the service according to the classification DCAT-AP of the European Union.
- Geolocation features of the service.
- The real-time feature of the service.
- The type of author of the application using the classification of reusers proposed in illustration 3 and described in section The reusers of data.
- If it had any mechanism of sustainability.
- If the services has a business model generating a revenue stream

Sustainability describes the economic viability of the service on mid-term perspective, either because there is an entity that supports the costs or because the service itself has mechanisms for generating income (business model). In example, in the case of a corporate service, the entity that publishes the service assumes its costs despite not having directly related income.

Just to make a comparison, the author has carried out a similar analysis for the open data ecosystem in Spain[6]. Figures for the universe of datasets reached 20026, quite similar to the found in the selected universe for this analysis, 20035 (See table 7). Conversely the amount of apps and services peaked 491, compare with the 209 found in this analysis (table 8).

6 This analysis will be published by the COTEC Foundation, (http://cotec.es) on September-October 2017

4 Exploratory analysis

4.1 Portals and infrastructure

According to the simplified maturity model for open data portals described in section 3 the results show three groups. First group whose members are those countries with a central open data portal with advanced features for data reuse. It includes countries scoring higher or equal to 90%, in the simplified maturity model(Brazil, Uruguay, Mexico, Panama and Paraguay).

Those portals run a DMS solution to manage their open data portals. All of them are running CKAN, and they have ready an API in order to grant automated access to the data. Besides this, all of them has a section to publish data-driven apps and services implemented based on the data. This apps section allows authors to get free promotion from the portal, and additionally is a demonstration for society of the potential impact of the released data. These portals sometimes are populated with instructions and guides on how to access the data with some code examples, etc. in order to speed up the development of new apps and services. The analysis of the created apps and services is discussed in section.

The second group comprises the rest of the listed countries scoring between 30 and 60 %. These portals lack of some of the critical features to nurture an ecosystem of organisations around them. For example lack of an API, or sometimes the solution is not a DMS, which makes mostly manual the systematic harvesting of data to be published. With the exception of Mexico none of them has an apps or services section where the developments based on the released data were shown.

Illustration 7: Datos abiertos Guatemala main page

Translation : Hello everybody www.transparencia.gob.gt will be available soon. Subscribe to our channel till the portal is ready. Retrieved 7-6-17.

The third group are those countries lacking of a proper portal, listing here not only Guatemala and Cuba, but also Venezuela, El Salvador, Nicaragua and Honduras. For Guatemala and Honduras, members of the OGP since 2011, their approach to the open government could be compromised because of the lack of a central repository to released open government data in order to be accountable and actually transparent.

In the case of Cuba and Venezuela there is not central repository of data but a web with statistical information.

4.2 Portal contents

According to the methodology described in section 3 a random sample was taken for the full portals considering a single universe of dataset population (see table 7).

In terms of licensing, although most of the sampled datasets were released with on open license[7] it is remarkable to find out some datasets released with a non-commercial licensing or even no license at all. It has to be noted that no explicit open license means, by default, a copyright license and therefore written consent to reuse the data even If they are available for download.

In terms of technical standards there are two main groups. First group comprises those portals releasing data using reusable private standards. The most popular format in this groups is Microsoft excel format. The second important group is populated by datasets released with an open standard with metadata, in example RDF, XML, KML, etc. Although just minor cases some examples were found releasing just pdf documents as data. From the point of view of reuse this can hardly be considered a dataset.

Access to the data is another analysis carried out. Findings include a significant amount of datasets published but not available to access (404 web access error). Possibly due to temporary non-availabilities. Further research should be done on this topic.

However the most popular method to access the data was the full download of the dataset. From the point of view of the reuser is far from being optimal, as long as it implies to filter the data in order to update, or to use the requested data. Finally another group includes those datasets available via API were some filtering is possible and access to specific content of a dataset is available. Further research should be done in terms of the documentation and liability of these resources.

The data model of the datasets, the meaning of the different columns of a group of data, is another interesting element to be analysed. Thus, most of the datasets publish the name of the fields but just a few an explanation of what the different fields represent, what are the valid ranges of values, and what type of field is included.

7 http://opendefinition.org

Table 6: Open data portals maturity

Country	Open data portal URL	Maturity
Brazil	http://dados.gov.br	100%
Uruguay	http://datos.gub.uy/	100%
Mexico	https://datos.gob.mx/	90%
Panama	http://www.datosabiertos.gob.pa	90%
Paraguay	https://www.datos.gov.py/	90%
Argentina	http://datos.gob.ar/	60%
Chile	http://datos.gob.cl/	60%
Colombia	https://www.datos.gov.co	60%
Costa Rica	http://datosabiertos.presidencia.go.cr/	60%
Ecuador	http://www.datosabiertos.gob.ec	60%
Dominican Republic	https://datos.gob.do/	60%
Peru	http://datosabiertos.gob.pe/	50%
Bolivia	https://datos.gob.bo/	40%
Cuba[8]	http://www.one.cu/	--
Guatemala[9]	http://datos.transparencia.gob.gt	--

Source: Own elaboration

Geolocation of data enriches the data released and allows further correlation with other data sources. Here the sample, although with a confidence interval of 18 points shows that the most popular is not to include geographic information at all. However the second most popular group comprises those datasets including not only coordinates but also some geographic text information (i.e. address, site identification) attached to the data.

Finally in terms of updating frequency, none of the sampled data were update with shorter frequency than once a day. This updating frequency

8 Cuba open data portal was in fact the national statistics office and therefore there were not an actual open data portal even though some data were available.

9 This link http://www.transparencia.gob.gt/?page_id=186 announces the ‚coming soon' for the open data portal for Guatemala (4-5-17)

restricts those reuses where real-time is an appreciated characteristic by the users.

Table 7: Number of datasets in Central open data portals

Country	Open Datasets in central portal
Argentina	69
Bolivia	24
Brazil	2.937
Chile	3.060
Colombia	4.079
Costa Rica	206
Cuba*	21
Ecuador	109
Guatemala	29
Mexico	8.786
Panama	44
Paraguay	188
Peru	310
Dominican Republic	31
Uruguay	142
TOTAL	**20.035**

* Cuban portal publication represents the categories of data available in the central statistics system

Source: Own elaboration. Retrieved May-June 2017

Globally all the datasets were qualified according to the reusability metric MELODA 4.13, and the results show that none of the sampled datasets could be qualified as optimum for reuse and that there are similar amount of datasets between these three categories, basic, good and deficient. Those datasets under the category deficient does not allow an effective reuse of the data and its publishing characteristics should be improved.

4.3 Portal apps and services

An exploratory analysis has been carried out on the data-driven services identified in the apps sections of the portals. This feature was only found in six out the analysed portals (Brazil, Mexico, Paraguay, Panama, Peru and Uruguay). The amount of listed services varies strongly as described in table 8.

Table 8: Number of apps / services for main central open data portal in Latin America

Country	Apps section of open data portal	# of apps / services
Brazil	http://dados.gov.br/paginas/aplicativos	11
México	https://datos.gob.mx/herramientas	42
Panama	http://www.datosabiertos.gob.pa/showcase	22
Paraguay	https://www.datos.gov.py/related	9
Peru	http://movil.softwarepublico.gob.pe/web/index.html	87
Uruguay	https://catalogodatos.gub.uy/related?page=1&type=&sort=created_desc	38
TOTAL		**209**

Source: Own elaboration. Sampled May-June 2017

The analysis includes 20 samples, close to the 10% of the universe population. The first remarkable result is that more than half of the sampled items led to 404 errors when trying to actually access to the service or application. Potential causes includes a poor maintenance of the services created or a dead rate of the services created beyond the control of the data portal maintainers. Further research should be done on this topic.

The topic of the applications were rather disperse across of the categories. Economy & Finance, Government and Public sector and Transport got more results than the rest, however due to the limited range of apps sampled no conclusive remarks can be made. The analysis of the contents reflects that the published services were transparency-driven.

In terms of the geolocation of the service, the sampled ones are divided in two similar groups. On one of them there are geolocated services and in another similar-sized group those non-geolocated.

However most of the created services do not provide services in real-time. Real-time in the sense that the service provided is not updated in real-

time. That is congruent with the datasets sampling where none of them has real-time characteristics.

In terms of authoring most of the developed applications were developed by the data publisher, but there are also examples of citizen developed services, professional reusers (for-profit) and academic reusers (see illustration 3 in section 2.5) .

The sustainability, namely somebody paying the maintenance of the application or service, was balanced between those having sustainability mechanisms and those which lack of it. But the clear difference is that none of the sampled applications were able create their own revenue stream and therefore has to be supported by the original entity.

5 Conclusions

The open government approach emphasizes the accountability of the public administrations. As a result of its implementation open government data should be publish regularly and profusely. The analysis has sampled the central open government open data portals, data published and data-driven apps and services across Latin America. The exploratory analysis carried out provides some hints about the real situation. Due to the reduced sample it is not possible to draw quantitative conclusions so qualitative ones will be provided instead.

On one hand the infrastructure for the data publication seems to be in place because most explored open data portals have a running portal with a small amount of exceptions. Additionally, DMS solutions are widely adopted, and in some cases automation of the data access is in place.

Open government data publication requires not only of an infrastructure, but also of an organisation supporting the data publication. Thus, it is found that some open data portals are profusely populated with data, while other are just pilot initiatives with some tens of datasets. Thus, although infrastructure is ready, the organisation maturity is diverse in order to include information on the open data portals.

Few conclusive remarks can be drawn from the publish data in terms of topic of the released data. The only common point is that information, even belonging to different categories are close to transparency efforts. Situation that it is congruent with an approach to data publication based not on the potential impact on the society for the data reuse but on the urgency for increased transparency coming from the citizens.

The previous remark about pilot projects in place remain in the analysis of the licensing of the data. Most of the sampled data were open licensed, however some non-commercial, or even some proprietary data are also published in open data portals. Similar impression arises from the analysis of the technical standards and access mechanisms. There is a mix of some advanced features for some datasets while other are in the most basic features. The geographic information of the data also reflects a mixed panorama of data. Majority lack of geographic information at all, while other include coordinates and additional field texts. Finally in terms of real-time approach to data publication, none of the sampled data could be considered a datajet, or stream of data.

Globally speaking analysed datasets appear in diverse reusability levels according to the metric MELODA but none of the in the advanced

reusability level and a significant percentage qualifies as deficient for reuse according to this metric.

The mixed impression of the data infrastructure and data publication cannot be maintained when analysing the ecosystem of data-driven services. The remarkable amount of abandoned services together with the absolute lack of business models to sustain the initiatives, draw a picture where it is plenty of room for improvement. Possibly this lack of sustainable business models could be related with the lack of data release in real time and with a proper geolocation of the data.

Summarising, this exploratory analysis has found a diverse maturity in terms of infrastructure and data publication but the early stages in term of the ecosystem of data reuse. A potential explanation for this situation is that the impulse based mostly on transparency demands does not meet the needs of the ecosystems of reusers.

5.1 Future and recommendations

As a recommendation, more attention should be paid to the fact that the released data could be reused with different purposes by a diverse group of reusers. Current impulse for data publication based on the demands of transparency requires of a complementary approach for making the data actually useful for reuse. Further results of this study could be reached by extending the sample to a bigger amount of data portals and with bigger sample size. Fortunately the rest of the world is plenty of good practices to learn from, and possibly the development of the area could be quick as long as the main infrastructures are ready.

Hopefully the development of data-driven services could boost a new economy based on the innovation and it will help to improve the lives of the population in this area.

BIBLIOGRAPHY

Abella, A., Ortiz-de-Urbina-Criado, M., De Pablos-Heredero, C. (2014). "Meloda , métrica para evaluar la reutilización de datos abiertos ". El profesional de la información , 2014, Nov-Dec, v. 23, n. 6.

Abella, A. (2017). Modelling of social and economic impact of open data in Smart cities' ecosystems. Doctoral dissertation. University Rey Juan Carlos Madrid. Escuela internacional de doctorado.

Abella, A.; Ortiz-de-Urbina-Criado, M.; De-Pablos-Heredero, C. (2016). The challenge of participation in public open data portals. Best practices, tools and recommendations. Congreso Gigapp, 2016.

Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2017). A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems. *Cities*, *64*, 47-53.

Carrara, W.; Nieuwenhuis, M.; Vollers, H. (2016). Open data in Europe 2016. Insights into the European state of play. European Commission. Directorate General for Communications Networks, Content and Technology.

Chan, Y.; Greenbaum, S I.; Thakor, A. V. (1986). "Information reusability competition and bank asset quality". Journal of banking and finance, v. 10, n. 2, June, pp. 243-253. http://goo.gl/JX7OHi http://dx.doi.org/10.1016/0378-4266(86)90008-7

Eaves, David (2010). "The three laws of open government data". Conference for Parliamentarians: balancing openness and the public interest in protecting information. http://bit.ly/eaves2010

European Union (2015). DCAT Application Profile for data portals in Europe. Version 1.1. Ref SC118DI07171. Available at https://joinup.ec.europa.eu/system/files/project/dcat-ap_version_1.1_0.pdf (June-2017)

Faniel, I. M.; Jocobsen, T. E. (2010). "Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data". Computer supported cooperative work, v. 19, n. 3-4, August, pp. 355-375. http://dx.doi.org/10.1007/s10606-010-9117-8

Jetzek, T., Avital, M. and Bjorn-Andersen, N.(2013) "Generating Value from Open Government Data". Thirty Fourth International Conference on Information Systems, Milan 2013.

Jetzek, T., Avital, M., Bjorn-Andersen, N. (2014a) "Data-driven innovation through open government data". Journal of Theoretical and Applied Electronic Commerce Research 9(2), 100-120.

Kim, H. (2005). Exploring data reusability: standardized representation of domain contents and feasibility testing. Doctoral dissertation, University of Minnesota, USA. ISBN: 0 542 31045 7

Kogan, N. (2014). Exploratory research on success factors and challenges of Smart City Projects. Thesis for the Degree of Master of Science. Department of Business Administration Graduate School. Kyung Hee University.February 2014. Retrieved from http://goo.gl/WPaVaI

Linders, D.; Wilson, S.C. 2011. "What is Open Government? One Year after the Directive," in The Proceedings of the 12th Conference on Digital Government Research (dg.o.2011), pp. 262-271.

Open Knowledge International. (2015). Open definition version 2.1. Available at http://opendefinition.org/

Ramirez-Alujas, A.V. and Dassen,N.; (2014). Vientos de cambio: El avance de las políticas de gobierno abierto en América Latina y el Caribe. Technical note. No. IDB-TN-629. Inter-American Development Bank. Available at http://bit.ly/ramirezalujas

Ramírez-Alujas, Á. (2014). Gobierno abierto. *EUNOMÍA. Revista en Cultura de la Legalidad*, 201-216.

Red.es (2017). Study of the infomediary sector 2016. Executive summary. Available at https://www.ontsi.red.es/ontsi/sites/ontsi/files/characterization%20of%20the%20spanish%20infomediary%20sector.%20Executive%20Summary.%20Ed.2016.pdf

Renear, A. H., Sacchi, S., and Wickett, K. M. (2010). Definitions of dataset in the scientific and technical literature. Proceedings of the American Society for Information Science and Technology, 47(1), 1-4.

Vickery, G; (2011); “Review of recent studies on PSI reuse and related market developments“. *Information Economics*. Paris

Zeleti, F. A., Ojo, A., & Curry, E. (2016). Exploring the economic value of open government data. *Government Information Quarterly*, *33*(3), 535-551.