A Model to Assess Open Government Data in Public Agencies

Mauricio Solar[1], Gastón Concha[2], Luis Meijueiro[3]

[1]Departamento de Informática, Universidad Técnica Federico Santa María, Chile
mauricio.solar@usm.cl
[2]Comisión Económica para América Latina y el Caribe (CEPAL), Chile
gaston.concha@gmail.com
[3]Fundación CTIC, Spain
luis.meijueiro@fundacionctic.org

Abstract. In this article a maturity model is proposed, named OD-MM (Open Data Maturity Model) to assess the commitment and capabilities of public agencies in pursuing the principles and practices of open data. The OD-MM model has a three level hierarchical structure, called **domains**, **sub-domains** and **critical variables**. Four capacity levels are defined for each of the 33 critical variables distributed in nine sub-domains in order to determine the organization maturity level. The model is a very valuable diagnosis tool for public services, given it shows all weaknesses and the way (a roadmap) to progress in the implementation of open data.

1 Introduction

The Electronic Government (e-Gov) development has been implemented from its beginnings with a model focused on services that the government provides. In recent years a change of model has been defined towards one focused on the citizen. This citizen-centric model is a new way of governing, and generates two implications in terms of public policies design and development of the digital strategy of the governments [1]:

- **New services model**: the services offered by the government must be designed and implemented focused on citizens (individuals or legal entities) as recipients of services rather than the government as a producer, who must incorporate essential attributes in the design with this new approach, such as: single window, multiple service channels, high usability standards, interoperability, and service levels defined as ex ante.
- **Open Government**: this concept includes participation, transparency and cooperation of citizens in public policies. In this area, *Open Government Data* (OGD) plays a significant role and has become a way to operationalize this approach.

The OGD concept is a work philosophy to empower citizens and provide them access and license to use the data generated by public entities, so that they can use,

store, redistribute and integrate them with other data sources. This data opening is justified both by encouraging citizen participation, strengthening democracy, as for being an innovation driving force by enabling the creation of new companies with these data.

The OGD concept is intertwined with the *Open Data* concept and *Linked Data* (linkeddata.org): *Linked Data* is a way of publishing data in such a way that it can facilitate the interaction between different data sources, while the concept of *Open Data* is oriented to a freely accessible data and without any restrictions at all to the people [2].

The *Open Data* approach attemps to put data at the disposal of all citizens, which has proven to generate an important public value. The authors of *Open Government* book [3] bring up three fundamental concepts for a better understanding of the *Open Data* impact:

- Public information is a kind of infrastructure, with the same importance level as other infrastructures (water, electricity, roads).
- Public value must be maximized as of existing data held by government.
- The open data magic is that it enables transparency and innovation.

Some countries have progressed beyond the mere access to data, where the Open Data Model has shown that it not only produces significant changes in the public sector, but it generates synergies in innovation and entrepreneurship. This is the case of the United Kingdom (UK) that through the organization The *National Archives* [4] has defined a data licenses model that allows the use and reuse of the information that is under the granted license. Citizens can access to public data, and can use it in commercial terms; this model allows the generation of entrepreneur in relation to these data.

Public Services (PS) of the government collect and produce large data volumes involving data as climatological, energy, economic, health, environment, agriculture, defense, public safety, social, cultural, budgets, among many others [5]. Unfortunately, these data may have restricted accesses, perhaps its existence is unknown, or can be in standardized or private formats, which brings into question some aspects like, why should I identify me to get public data? Why should the request for public data be justified? Why is it that a software sometimes expensive to be processed (such as SPSS), or private as Excel must be bought? How to make citizens and civil society to taking full advantage of this data? What kind of services should governments provide so as to increase citizen's participation in OGD initiatives?

Considering the importance and relevance of implementing Open Data in PS, a proposal of the model for diagnosing PS capabilities is presented in this paper to develop Open Data. The model, called Open Data Maturity Model (OD-MM) is based on the principal elements found in literature, described in next section. The elements of the model are described in section 3, and development methodology and validation in conceptual terms is presented in section 4. Conclusions can be found in last section.

2   Alternatives for diagnosing Open Data Implementation

Although it was not dully made official, a classification proposal exists about how open and usable are the data that a public agency can provide. Sir Tim Berners-Lee or TimBL (www.w3.org/DesignIssues/LinkedData.html) proposed the well known "five stars" test for the data publication.

In December 2007 a group of researchers in Sebastopol, California, developed the 8 principles for OGD's, designed for a better understanding of these concepts and to define why OGD is essential for democracy. One of the key concepts for the development of these principles was to establish that the information becomes more appreciated if shared, because it not only benefits the end user which is civil society, but it also makes more efficient the use of public resources. The 8 principles established for considering government data as open are found in www.opengovdata.org/home/8principles.

Reggy in [6] defines a model with 4 levels for each of the 8 principles above mentioned, based principally on the W3C guidelines and the COI of the UK. For each principle a score is assigned according to the following: Level 0 = 0%; Level 1 = 33%; Level 2 = 66%; and Level 3 = 100%. A compounded indicator allows measuring the overall quality of each program evaluated by simply calculating the media of the score associated to the 8 principles.

MELODA (MEthodology for reLeasing Open Data) is a tool to accelerate releasing of information to society. It was not designed exclusively for public agencies, but it also included the private sphere, weakening its proposal as a tool of diagnosis for the public sector (gobernamos.com). MELODA evaluates available information from a data source (the same information that anyone could reach, including commercial uses mixed with private sources). This requirement restricts the number of analysis dimensions, which currently covers three dimensions: Legal Frame, Technical Standards for data releasing, and Accessibility to information. Five maturity levels are considered for each dimension. Levels from 1 through 5 for each dimension are marked as 0%, 25%, 50%, 75% and 100%. Each dimension is weighted the same as the others, so the global assessment for a single data source is composed by one third of each assessment. The global assessment of a PS that releases data is the average of each data source assessment released by the institution.

The *Carter Center* (www.cartercenter.org/documents/2012.pdf), in its 2008 Action Plan, provides policies focused on access to information as a human right in all cultures and government systems. It recommends that governments and international organizations must ensure this right by providing: fair exercise of the access rights, training of public officials in the practice and application of access rights, public education to empower such a use of the law, among others.

The stage model for OGD proposed in [7] has two main dimensions, namely organizational and technological complexity and added value for data consumers.

According to literature, OGD presents challenges in several fronts, the most relevant are:

- Experience in different countries shows that it is necessary to train those who will be responsible for OGD in each public agency, which takes time.
- Not all data have the same quality; how reliable they are, how they are represented, and so on. It is necessary to establish a set of metrics that could help consumers of these data.
- It requires a strong political support and to considering an OGD as a fundamental policy to improve the government transparency. Without an appraisal from the political world, every effort will remain in good intentions. Moreover, this support should be reflected in a resources allocation, since as any other public policy, implementing OGD requires time and money in order to make it possible to managers to carry it out (as described in [8]).

Important elements that can be identified in literature and that should be considered when diagnosing the implementation of Open Data at PS level are those that stand out in successful cases described in literature. Among these perspectives the following are important to be considered:

- The **establishment** of a PS, given that the importance of leadership and strategy in OGD initiatives is highlighted in literature.
- The **legal aspect**, allows to having a legal frame when implementing *Open Data*.
- The **technological perspective** as for the accomplishment of *Open Data* principles, such as access to data, data quality and its availability.
- The **citizen perspective** as from participation and collaboration point of view.
- And **developers** and entrepreneurs in the reuse of data.

All these elements are considered in the maturity model proposal to assess the capabilities and maturity of public institutions in the *Open Data* implementation. We merge the first two perspectives (establishment and legal aspects) into a single domain, as well as the last two above mentioned (citizen and developers perspective), so that the proposed model has three perspectives that are detailed in next section.

3   Basis of the OD-MM Model

It is important to clarify that the proposed model is intended to be a reference to diagnose the capacity to face the OGD in Public Services (PS). It does not intend to be a model to establish specific processes required for better services delivery but a reference for the design, management, monitoring and performance control of these processes adapted to the reality of each organization and the particularities of the public administration.

The model is based on the definition of a set of 3 elements organized in a hierarchical structure. Their key elements are the "**Domains**", "**Sub-domains**" (**SD**) and "**Critical Variables**", which can be evaluated with respect to OGD. The "domains" are SDs logical associations which in practice have to mature and therefore, are subject to evaluation.

Domains are the heart of the model because they are set on different Capacity Levels (CL's) that contrast with those available in a specific organization. Three domains and 9 SDs were defined in total (3 per domain). In total there are 33 critical variables distributed in the 9 SDs. They are hereunder listed.

3.1 ELP Domain: Establishment and Legal Perspective

This domain determines the organization's ability to articulate a consistent vision of OGD. That is, the existence of an IT strategy aligned with a business strategy and that explicitly considers a decision and vision of the organization to be incorporated to open government. It involves high-level activities that allow to managing all resources according to the business vision and strategy and its priorities, including alignment with the guidelines of the national e-Gov, in particular OGD; includes the level of existence and confrmation of internal laws and regulations that facilitate the implementation of OGD policies and activities, and the ability of an entity to accurately perform organization, management and training activities as planned. Its SDs are hereunder described.

Subdomain ELP-1: Strategy, Leadership and Establishment
Includes high-level activities that allow managing all resources according to the business vision and strategy, and its priorities; it includes the alignment with the OGD guidelines at a national level. Its CL is determined by the following **variables:**

- *Strategy:* There is an OGD strategy aligned with the central level strategy.
- *Leadership:* Capabilities are available to successfully lead an OGD process.
- *Establishment:* Ability to have an organization to properly managing OGD programs with all necessary coordinations with other agencies.

Subdomain ELP-2: Laws and Regulations
It comprises the existence and verification level of internal laws and regulations that facilitate the implementation of the OGD policies and activities. Its CL is determined by the **variables**:

- *External Regulations:* Ability to comply with external regulations.
- *Internal Regulations:* Ability to comply with internal regulations.
- *Licenses:* Ability to accomplish the general regulation on licensings and generate eventual specific licensings of data sets generated by the entity without breaking with open data principles.

Subdomain ELP-3: Management
This subdomain includes the ability of an entity to appropriately perform organization, training and management activities as planned. Its CL is determined by the following **variables**:

- *Training:* Ability to provide staff trained in OGD suitable skills.
- *Project Management:* Ability to managing OGD projects, according to standard procedures and incorporate Open Data principles on related projects.

- *Performance Assessment:* Includes the development of metrics and measurement of periodical results about the initiatives and programs associated to OGD and ensure an appropriate internal and external coordination.

3.2 TPE Domain: Technological Perspective

This domain establishes the technological capacity of the organization to articulate a consistent vision of open government. That is, the existence of a technological strategy aligned with the strategy that explicitly considers the OGD best practices in the organization to be incorporated to the open government. Its SDs are hereunder described.

Subdomain TPE-1: Safety and Availability
Involves those activities related to ensuring the existence of protocols and mechanisms to protect data infrastructure from external and internal attacks; the idea is to ensure the availability of data at all times. Its CL is determined by the **variables:**

- *Safety Systems:* The organization has safety systems that protect data integrity in the context of a systematic implementation plan.
- *Data Availability*: Data should be available in time where their access can provide them an added value, i.e., in a timely manner.
- *Data Updating:* Updating protocols, manual or automatic.
- *Tools for Measuring the Level of Use:* Control of the data use level, i.e., the existence of data collecting mechanisms of data use level.

Subdomain TPE-2: Access
Involves the activities related to the portals enabling, the existence of *Data Sets*, etc. Its CL is determined by the following **variables**:

- *Automated Data Reading:* Information is accessible through database language or other ways of access.
- *Metadata:* Existence of metadata and a type of them.
- *Categorization and Discovery Facilities:* Easiness for search, creation, and maintenance in the Web portal of taxonomies.
- *Use of Semantic Technologies:* The Web portal provides support for the semantic enrichment of data sets.

Subdomain TPE-3: Data Quality
Involves the activities related to maintaining and managing the organization's data quality, in terms of Open Data existing definitions and level of interoperability of the organization's data. Its CL is determined by the following **variables:**

- *Data Format*: Use of non-proprietary formats that facilitate interoperability.
- *Free Data*: The data are not restricted by copyright or other legal restrictions that could limit their reuse.
- *Primary Data:* It should be primary data, avoiding secondary processed data.
- *Data Completeness:* Data must be complete (all the aspects are given).

3.3   CEP Domain: Citizen and Entrepreneurial Perspective

This domain establishes the organization's ability to listen to public opinion, to involve citizens and collaborate with developers of applications that improve the transparency of the organization. Its SDs are hereunder described.

Subdomain CEP-1: Data Re-use
Open data publication status. Its CL is determined by means of **variables**:

- *Open Data Developed Initiatives:* Number of Open Data projects completed or in progress, and status of them.
- *Number of Open Data Available:* Volume (ratio) of OD data sets published by the entity in relation to the global information provided by various means.
- *Single Access Point*: Availability of access catalogs to data sets.
- *Data Access Measurement:* Existence and management of access indicators and/or downloading of data sets, analysis of results and proposal of improvement measures (selection criteria, promotion, etc.).

Subdomain CEP-2: Developers
Degree of involvement in the encouragement and assistance to the work of reusers agents (RISP: Reusers of public information). Its CL is determined by the **variables**:

- *Data Gratuitousness*: Free data access and/or data downloading.
- *Reuse Encouragement:* Existence of aid resources (documents and material published or promoted by the institution) as well as promoting, formative and networing activities, conducted or promoted by the entity.
- *Complains and Conflicts Resolution:* Reception, recording and resolution of complains and conflicts in re-using matters.
- *RISP Project Financing:* Use of available fundings for the development of reusing applications, whether it is financing offered or promoted by the entity itself or others, but advertised/processed by it.

Subdomain CEP-3: Participation and Collaboration
It includes a certain level of listening-in and adaptation to citizen's demands, as well as to dialogue. Its CL is determined by the following **variables:**

- *Participation and Collaboration Means:* Communication channels exist between citizens and the entity about RISP matters, and type of collaboration developed.
- *Participative Transparency:* Management and advertising of citizen's participation.
- *Active Listening:* Implication's degree of entities in public participation and reaction when facing the improvement to data publication and reuse.
- *Data Use Measurement*: Presence and management of indicators of use and/or applications' demand based on RISP.

3.4 Capacity and Maturity Model

Capacity is a property of each SD. The ability of a SD is determined based on the Capacity Level (CL) of its Critical Variables, i.e. what is actually measured is the ability of Critical Variables to meet certain requirements, then these capacities are weighted according to their importance and the result of this weighting is the final CL of the SD.

The Maturity Level (ML) instead, is a property of the organization as a whole. Each ML will correspond to a SD default setting in predefined CL. The ML prescribes a "roadmap" or path of improvement for the organization.

One of the important objectives of the developed model is that it can produce progressive evolution alternatives of capabilities and maturity. In order to achieve this, the generic model to be used to define the CL characteristics for each model variable is here described. Then, the relationship between the variables' capacities and their respective SD is as well described, and finally how to determine the maturity of the organization once their SD capacity is known.

For each SD an incremental measurement scale exists based on a score from 1 to 4. This scale is associated with a generic qualitative capacity model described below.

Level 1: Inexistent Capacities

- Capabilities do not exist or the SD is approached in an ad-hoc and reactive manner, tends to be applied on an individual case by case way.
- There is evidence that the SD's were recognized and need to be approached.

Level 2: Emerging Capacities (unformal)

- An intuitive regular pattern to approach the SD's is followed. Different people follow similar procedures to approach the same task.
- There is no formal training or divulgation of procedures, and responsibility to follow them up rests on each individual.

Level 3: Existent Capacities

- The procedures related to the SD's were defined, documented and communicated.
- There is a formal training to support specific initiatives related to SD.
- Procedures are not sophisticated; they rather are the formalization of existing practices.
- Monitoring and measuring of compliance with procedures is possible, as well as taking actions when the apparent SD's do not effectively work.
- Standards and guidelines established apply throughout the whole organization.

Level 4: Advanced Capacities

- Procedures have reached the level of best practices and continuous improvement is applied.
- The use of market standard or world-class tools helps to optimizing the SD's.

Relation of Variables, Capacities and SD

Capacity is a property of each SD and is obtained by measuring the CL of its critical Variables (V_i). Whereas there should be a direct relationship between both capacities it has been chosen as a calculating mechanism the arithmetic average of the SD constituent variables capacities. However, considering that for a given *Open Data* strategy or that for a level of development for a given country there are more relevant variables than others, a set of ponderers for each group of variables has been defined. Thus the CL of a SD turns out to be a weighted sum (w_i) of their constituent CL variables (Eq. 1). Table 1 show weighting values initially defined to be used in the pilot stage of the model.

$$CL_{SD} = \sum_{i=1}^{n} \left(CL(V_i) \times w_i \right) \qquad \text{(Eq. 1)}$$

Table 1. SD set of variables on each domain

Domain	Subdomain	Variables	Weight
Establishment and Legal Perspective	Strategy, Leadership and Establishment	*Strategy*	*30%*
		Leadership	*40%*
		Establishment	*30%*
	Laws and Regulations	*External Regulations*	*20%*
		Internal Regulations	*40%*
		Licenses	*40%*
	Management	*Training*	*30%*
		Project Management	*30%*
		Performance Assessment	*40%*
Technological Perspective	Safety and Availability	*Safety Systems*	*20%*
		Data Availability	*30%*
		Data Updating	*30%*
		Tools for Measuring the Level of Use	*20%*
	Access	*Automated Data Reading*	*30%*
		Metadata	*30%*
		Categorization and Discovery Facilities	*20%*
		Use of Semantic Technologies	*20%*
	Data Quality	*Data Format*	*30%*
		Free Data	*25%*
		Primary Data	*25%*
		Data Completeness	*20%*
Citizen and Entrepreneurial Perspective	Data Reuse	*Open Data Developed Initiatives*	*30%*
		Number of Open Data Available	*30%*
		Single Access Point	*20%*
		Data Access Measurement	*20%*
	Developers	*Data Gratuitousness*	*20%*
		Reuse Encouragement	*40%*
		Complains and Conflicts Resolution	*20%*
		RISP Project Financing	*20%*
	Participation and Collaboration	*Participation and Collaboration Means*	*30%*
		Participative Transparency	*20%*
		Active Listening	*30%*
		Data Use Measurement (Applications)	*20%*

The ML is determined by a set of values for the SD model, as shown in Table 2. The advantage of this scheme is its flexibility since it only establishes a minimum set of SDs, important in a given ML. The remaining SDs are left to the PS criterion.

Table 2. Template of organizational maturity based on a set of priority SDs

Domain	SD	ML 1	ML 2	ML 3	ML 4
Establishment and Legal Perspective	Strategy, Leadership and Establishment			2	3
	Laws and Regulations			3	4
	Management		2	3	4
Technological Perspective	Safety and Availability			2	3
	Access		2	3	4
	Data Quality			2	3
Citizen and Entrepreneurial Perspective	Data Reuse		2	3	4
	Developers		2	3	4
	Participation and Collaboration		2	3	4

4 OD-MM Model Development and Validation Methodology

The design methodology included to studying papers and published articles, study committee reports, evaluations and gray literature (see references), as well as telephone interviews, *skype* meetings and teleconferences.

The work team contacted various groups of principal stakeholders in *Open Data* for interviewing and using them as primary sources of information.

Interviews were conducted in two rounds. The first round was focused on 9 experts that had been directly or indirectly involved in the development of the Open Data projects, in the Government of Chile, Colombia and in the state of NSW in Australia. Information collected in this process was gathered, and from literature review, questions were structured for the second round of interviews. The second round of interviews was focused on 4 experts with a more extensive domain and including people with experience in the region, Latin American governments and representatives of foundations and civil society defenders of the Open Data movement.

4.1 Determination of the Weight of Variables by SD

Determination of the weight of variables must follow a methodology in which at least three experts should be involved: a government representative, a citizen representative and/or *Open Data* applications developers, and at least one *Open Data* expert independent of the first ones. The methodology used in this study consisted of 3 rounds:

- **First round**: Each expert assigns independently a weight to the variables according to their criterion within the SD. In each SD the 100% is distributed among the variables that comprise the SD.
- **Second round**: the weight of each variable proposed given by each expert in all three domains is compared. To those variables proposed by the experts with equal weight, that weight is assigned.
- **Third round**: in cases where any difference exists between proposed weights, a negotiation starts, and is repeated for each SD. For weights with no big difference between the proposed ones, the intention is to come to a consensus value. If there is no consensus, each expert justifies his proposal and relative importance in SD. Finally, if no consensus is reached, it is submitted to the opinion of an external expert whose last word will be accepted without discussion.

4.2 Validation of OD-MM Model

To validate the OD-MM model, this was sent to various experts related with the OGD topic in Latin America. Representatives of the Government of Chile and Colombia, civil society and developers of Open Data applications made their comments which after being analyzed lead to modifying those that allowed to improving the proposal of model design to validate it.

With the validated model, in conceptual terms, a Web tool was implemented which allowed the data colletion required by the model for its validation by means of a pilot in-situ. For this reason, a methodology was established to apply this pilot. The URL of the Web tool is http://odmm.inf.santiago.usm.cl/ (interface in Spanish).

For application of the pilot, experts from Chile, Colombia, Dominican Republic and El Salvador were contacted. In these countries a sample of 10 public institutions were selected, based on previously defined criteria. Results of the pilot will be reported in a future publication, which for space availability reasons are not possible to describe in this article.

5 Conclusions

A model was designed to properly meet specificities of the public sector at a regional level, being a frame of reference to identify the areas that support the OGD strategy, based on OGD international best practices.

The resulting model incorporates a hierarchical structure according to the relevance of the analyzed domain. The structure includes 3 domains, 9 sub-domains and 33 variables. This structure helps to a better adjusting to the diverse reality of PS in their preparation for joining the Open Government.

This model distinguishes between capacity, as a characteristic of a SD, and maturity as the organization's property as a whole. The capacities of variables contribute in a weighted manner to the SD capacity generating another adaptability element.

Domains on which the ML of capabilities in each PS is established, enables a continuous improvement and, therefore, a continuous progress towards higher levels of maturity of each organization. For space limitations, it was not possible to show the roadmap generation in this article.

Consequently, given the adjustability attributes and incorporation of a cycle of continuous improvement to which the grounds of the model are closely related to elements considered as the best international practices, we believe that the implementation of the model at a regional level will be a powerful diagnosis tool for Open Data.

References

1. Barros, A., 2011. Open Data: new paradigm in data management. *Revista BITS* de Ciencia. Computer Science Dept., University of Chile. Num. 6. www.dcc.uchile.cl/revista (in Spanish).
2. Sifaqui C., Otto O., Almazán F. and Hernandez D., 2011. On the road to the Semantic Web: experiences in the Biblioteca del Congreso Nacional de Chile (BCN). *Revista BITS* de Ciencia. Computer Science Dept., University of Chile. Num. 6. www.dcc.uchile.cl/revista (in Spanish).
3. Lathrup D. and Rume L., 2010. Open Government: Collaboration, Transparency, and Participation in Practice. O'Really Media, 432p.
4. The National Archive, 2012. www.nationalarchives.gov.uk/doc/open-government-licence
5. WEF, 2011. The Future of the Government.
6. Reggy, L., 2011. Benchmarking Open Data Availability across Europe: The Case of EU Structural Funds. European Journal of ePractice, www.epracticejournal.eu Nº 12.
7. Kalampokis E., Tambouris E and Tarabanis K., 2011. Open Government Data: A Stage Model. In M. Janssen et al. (Eds.): EGOV 2011, LNCS 6846, pp. 235-246, 2011.
8. McDermott P. 2010. Building open government. Government Information Quarterly 27, pp. 401-413.