Open Government Data: Towards a comparison of Data Lifecycle models

Hanae Elmekki
Université Mohammed V- Ecole Mohammadia d'Ingénieurs (EMI)
Rabat
Morocco
hanae.elmekki@gmail.com

Dalila Chiadmi
Université Mohammed V- Ecole Mohammadia d'Ingénieurs (EMI)
Rabat
Morocco
chiadmi@emi.ac.ma

Hind Lamharhar
Université Mohammed V- Ecole Mohammadia d'Ingénieurs (EMI)
Rabat
Morocco
hd.lamharhar@gmail.com

ABSTRACT

Government, through Open Government Data "OGD, becomes one of the important producers of open data. OGD is an opportunity to create valuable services and innovative products useful for citizens as a primarily targeted consumer. However, the expected benefits of OGD are not yet met. That is to say, several research communities' studies insist on the necessity of creating valuable data in order to generate valuable services. These studies are still insufficient for a shared understanding of how OGD contribute to the creation of value. For this purpose, this paper presents a review of a set of data lifecycle models compared against their contribution to the creation of value in the context of OGD.

CCS CONCEPTS

• **Information systems ~ Data management systems**

KEYWORDS

Open Government Data, data lifecycle, data value creation.

ACM Reference Format

H. Elmekki, D. Chiadmi, H. Lamharhar 2019. Open Government Data: Towards a comparison of Data Lifecycle models. In Proceedings of 6th Annual International Conference on Arab Women in Computing and Technology, Rabat Morocco, March 7-9 2019 (ArabWIC 2019), 6 pages.

ArabWIC 2019, March 7–9, 2019, Rabat, Morocco

ACM ISBN 978-1-4503-6089-0/19/03..$15.00
https://doi.org/10.1145/3333165.3333180

1. Introduction

The openness of government data was guided, in its early age around the 60s and 70s, by a political vision with the objective of promoting transparency and democracy **[1]**. In fact, many local and national governments have embraced the Open Government Data (OGD) initiative by making the data access free to every citizen.

Based on the existing literature **[6][27]**, OGD's initial purpose was to ensure the welfare of citizens and improve

the relationship between them and the government by enforcing transparency, promoting efficient and effective services delivery to the public, and involving citizens in decision-making**.**

Furthermore, making government data open, accessible, and reusable is an opportunity to create employment and promote scientific innovation **[2]**. These benefits have encouraged massive exploitation (or use) of ODG in different domains, for example, the use of OGD in economic field **[19]**, smart cities **[21]**, and public management **[20]**.

In stating this, we note that the use of OGD can have a positive impact on different areas and contribute to the creation of different types of value that we describe as follows:

- **Politic value** is accomplished through strengthening transparency and democracy and sharing public sector information and facilitating access as well as the use of open governmental data;
- **Economic value** is achieved through promoting technology innovation and job creation and also enabling collaboration between government and private institutes to create new opportunities for employment;
- **Social value** is fulfilled by encouraging citizens to participate more and engage in decision making alongside the government rather than passively consuming public services.

To create politic, economic and social values mentioned above, OGD defines a process chain to ingest inputs (e.g. Raw data), and apply specific actions in order to generate

a relevant value (eg. Information/knowledge). The process chain is called a data lifecycle "DLC" and are vital to achieving the sole purpose of OGD which is to bring both the government and citizens closer.

A DLC model design is targeted for different use cases **[18][26][17].** It can be used in different domains and industries including technologies such as Cloud Computing **[22]**, serving the interest of companies **[28]** and governments for managing their data **[6][14][16][23][25]**. In this paper, we aim to present and to compare a set of DLC models. These models have been developed either in the context of government field or in the context of OGD. We highlight particularly how they contribute to the creation of OGD values.

The paper is structured as follows. The first section is an introduction. The second section introduces OGD and describes the process of the creation of value. Then the third section gives an overview of comparative studies for DLC in literature. After that, we compare in the fourth section some DLC models developed in governmental context with the respect to the creation of OGD's values. The last section concludes the paper.

2. Open Government Data and data creation value

In this section, we provide different definitions of Open Government Data concept, then we give an overview of the existing literature about data value creation.

2.1 Open Government Data

In the context of our study, OGD represents a category of data placed online by public administrations for free access and exploitation.

From our literature review, we identify that OGD is described in different ways. It is defined as an initiative, a policy or a program in **[4] [5]**, a system, a technology or a platform in **[6][7]**, and as a philosophy or a strategy in **[8] [9]**. The reason behind this divergence refers to the various perspectives from which the OGD is discussed. In the political field, OGD is considered as an initiative launched by governments to promote public services **[10]**. From a business perspective, OGD is considered as a strategy and the new business model to follow **[1]**. From the software developer view, OGD is seen as an opportunity for innovation and developing new platforms and technologies **[11]**.

A set of OGD principles was given by the Open Government Working Group **[12]** and extended by the Sunlight Foundation **[13]:**

- Completeness: data must be fully published with necessary metadata;
- Primacy: maintain the degree of granularity of data as collected from the source;
- Timeliness: data should be always up-to-date;
- Ease of Physical and Electronic Access: data should be easily accessible;
- Machine readability: data must be understandable and easily processed by machines;
- Non-discrimination: data should be accessible without authentication controls or registration requirements;
- Commonly owned or open standards: data must be provided in accessible formats and for free;
- Licensing: data cannot be copyrighted but it's possible to create an appropriate open license to allow free access to data;
- Permanence: data should be available any time;
- Usage Costs: the cost to access data must be minimal.

Achieving those OGD' s principles leads to reach the objectives that are envisioned from the initiation of OGD. Those objectives (cf. fig 1) can attain the political field, the social field, and the economic field **[16] [27]**.

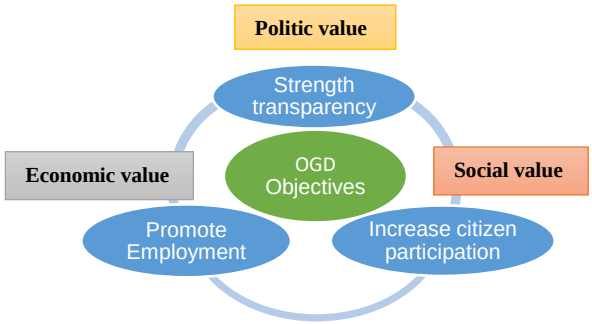

Figure 1: Values created through the OGD initiative

In the next section, we present how the creation of data value has been addressed in literature.

2.2 Data creation value

As we explained above, a data lifecycle DLC is required to create politic, economic and social values. From a conceptual point of view, a data lifecycle can be seen as a data management process including procedures and best practices which allow faster results and dependable long-term references **[3]**. It has recently been proposed as an effective data management solution that facilitates the organization of data and the extraction of knowledge in complex data systems. The following is a list of some recognized definitions:

- DLC models define the sequence of phases in the data's life, specify the management policies for each phase, and describe the relationship among phases **[14]**.
- DLC models describe the steps followed to handle data from the provider to the consumer and it does not stop after publication, otherwise the data would be lost **[3]**.
- Throughout their life cycle, data acquire a value that is equivalent to "making the data more usable in general or specific use cases" **[15]**.

The steps of creating value upon data constitute a Data Value chain that is a part of the data lifecycle. The concept of "value chain" is originally from the industry field to identify where, when and how value is created within

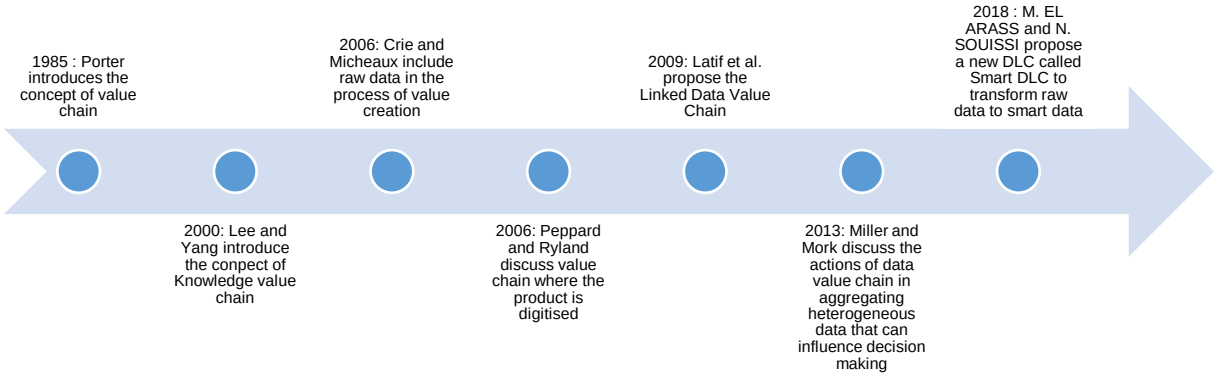

Figure *2*: The chronology of the data value chain concept

activities **[1]**. Authors in **[16]** consider the data value chain as the heart of knowledge economy that aims to use data to develop traditional sectors such as transport and health. The chronology of evolution **[15]** (cf. fig 2) of the data value chain concept shows that it has gone through several levels: knowledge creation, linked data creation, data collection and integration from heterogeneous sources, and smart data creation.

In fact, creating value means providing data that are usable to create added-value services and help in decision-making hence the importance of the DLC in managing OGD and creating value.

The next section presents an overview of comparative studies for DLC models in the literature.

3. Literature review

We identify various characteristics used to compare DLC models.

Authors in **[29]** took an interest in the DLC for Big data. They carried out a comparison of 17 DLC models with respect to the six challenges of BIG DATA including value, volume, variety, velocity, variability, veracity. This study emphasized the importance of the inclusion of "data quality" and "data security" in the veracity challenge during the design of a DLC model. It stated also the lack of a generic model designed independently from the field of study.

Authors in **[18]** came on the same conclusion. They compared 4 DLC models relating to both their sequential stages (e.g. Create, Publish/Share) and their complementary aspects (e.g. metadata, security and quality) with the aim to verify if those models explain how each phase is made and how data are used. However, we notice that those studies exclude from their evaluation important key factors in the process of the creation of valuable data including data interoperability and user feedback.

In the next section, we review some DLC models developed in the governmental context. After that, we will compare those models with respect to OGD's values' creation.

4. DLC comparison and discussion

Note that to achieve the objectives expected from OGD, data need to be managed efficiently which is the main goal of the development of DLC models. There are in literature, several DLCs designed for the government context. Some of them are developed in the specific context of OGD. For our study, 5 DLCs were selected and organized into two groups (cf. Table 1). All of them aim to create an architecture to present the life of data from the phase of creation (collection) to the step of consumption.

Table 1. DLC selected

DLC designed in the context of	
Gov Data	**Open Gov Data**
Ku and Gil-Garcia DLC overviewed	Government Data Life Cycle
Data value spectrum	Data value Network DVN
	Comprehensive Scenario Agnostic Data Life Cycle (COSA-DLC)

In this section, we will compare those selected models based on their contribution to the creation of OGD's values. Firstly, we outline the criteria of analysis, then we describe each model, and finally provide the comparison results.

We selected 3 criteria that are most closely associated to the creation value:

1. Contribution to the creation of politic value: the politic value results from opening public governmental information as much as possible in order to enhance transparency and root democracy. The fact of publishing OGD is insufficient to create politic value because data need to be transformed to information/knowledge for better decision-making.
2. Contribution to the creation of social value: the social value is fulfilled by encouraging citizens to collaborate with the government in decision-making. Through this collaboration, citizens could express their interest and feedback concerning governmental services. More specifically feedback can be helpful to improve the data published by the government.

3. Contribution to the creation of economic value: the economic value is achieved through promoting technology innovation and also enabling collaboration between government and private institutes to create new employment opportunities. In fact, the creation of economic value depends on the quality and interoperability of published OGD.

4.1 Ku and Gil-Garcia DLC overviewed

Data is essential to the development of smart cities through establishing data-driven decision-making approach **[23]**. Inspired from the data life cycle Ku and Gil-Garcia **[24]**, **[23]** propose a model to respond to the challenges facing city government in the context of smart cities. They argue that (i) governments should focus on organizational, technical factors and data quality and (ii) integrate those dimensions on the first stages (data acquisition and storing) of the data life cycle to provide valuable data able to provide public value.

Our assessment: The model ignores the data interoperability issue, the user feedback, and "Data Transformation" layer. Therefore, it didn't contribute to the creation of politic, social and economic values expected from OGD.

4.2 Data value spectrum

Through six research projects that consume big data to design Intensive Information System in several fields including automobile, transportation, healthcare, wellness, telecommunications from industry and governments field, "Data-value spectrum"**[6]** is developed as a data value chain constituted of nine factors (Data source; Data collection; Data; Data analysis; Information on the data source; Information delivery; Customer (information user);Value in information use; Provider network) to manage data in the case of Intensive Information System.

Our assessment: in this model, the step of "Data Transformation" and the aspects of data quality, user feedback, and data interoperability are missing. Therefore, it didn't contribute to the creation of politic, social and economic values expected from OGD.

4.3 Government Data Life Cycle

Designed in the context of OGD, this DLC model **[16]** defines the processes (activities) of the life cycle and the actions executed during each process. The model involves the implication of each stakeholder.

Our assessment: This model offers the phases of publication and transformation that help both to the generation of the politic value. However, data quality, user feedback, and data interoperability are missing in this model. Therefore, it fails to contribute to the creation of social and economic values expected from OGD.

4.4 Data Value Network

Data Value Network DVN **[25]** was developed in the context of OGD as a set of independent activities having the aim of creating value upon data as a non-tangible product in order to exploit it as a product. It consists of independent activities. Each activity is composed of a number of actions and each action can be considered as a data value chain. As each action needs a set of processes to be executed that constitute a nested value chain. The advantages of DVN is non-sequential, supporting multiple actors processing in independent activities.

Our assessment: This model is the extension of the model Government Data Life Cycle **[16]** described above. It provides the stage of "Data Publication" and "Data Transformation" as key factors to the generation of the politic value. The concepts of user feedback and data interoperability are not considered. It includes "Data Quality" as a complementary phase to the process of value creation. However, the list of data quality factor is not specified. Therefore, it didn't contribute to the creation of social and economic values expected from OGD.

4.5 Life Cycle (COSA-DLC)

In the context of Big Data, the model Comprehensive Scenario Agnostic Data Life Cycle (COSA-DLC) **[14]** is developed to cover all phases of data management from data acquisition to data preservation and processing intended to provide higher data value. This model includes also other fundamental aspects related to data quality and data security. Developed to addresses the 6Vs challenges (namely Value, Volume, Variety, Velocity, Variability, and Veracity) of Big Data, each block of the COSA-DLC model is developed in a set of phases adaptable and customizable to the field of study.

Our assessment: This model ignores the concept of user feedback and data interoperability and the phase of "Data Publication". It includes "Data Quality" as a sequential phase to the process of value creation without the specification of the list of data quality factors. Therefore, it didn't contribute to the creation of politic, social and economic values expected from OGD.

5. Discussion

We conclude that all the 5 DLC models have the common objective of providing a data management architecture covered as much as possible the different stages of data life based on how detailed their description is. Table 2 shows the results of the DLC models comparison, based on their contribution to the creation of values expected from the OGD initiative.

Table 2. Summarized comparison between DLC models

	DLC model	Politic value		Economic value		Social value
		Data Publication phase	Data Transformation phase	Data Quality	Data Interoperability	User feedback
Open Gov Data	Government Data Life Cycle	✓	✓	×	×	×
	Data value network DVN	✓	✓	✓	×	×
	COSA-DLC	×	✓	✓	×	×
Gov Data	Ku and Gil-Garcia DLC overviewed	✓	×	✓	×	×
	Data value spectrum	✓	×	×	×	×

Each model is marked (✓) if it addresses the corresponding criteria, otherwise it is marked (✖)

From this table, we observe that most discussed models include the "Data Publication" phase as a fundamental step to achieve politic value expected from OGD. However, the phase of "Data Transformation" is excluded from the models "Ku and Gil-Garcia DLC overviewed" and "Data value spectrum", for that they were marked "✖" in the table and then they fail in the process of the creation of politic value comparing to other models. Furthermore, the table shows that "Data Quality" is included in most models as a supplementary or a complementary phase. Although, the "Data Interoperability", as a key factor to the creation of economic value is not considered in any model. For this reason, we have marked "✖" for all models. Finally, we note that the discussed models do not show any focus, among their lifecycle steps, on the aspect of "User Feedback", as a key factor in the creation of social value.

Consequently, based on this study we deduce that there is still a lack of DLC models that enable us to take full advantage of the OGD initiative.

6. Conclusion

This paper discussed data lifecycle DLC as an approach for managing data since the time of their creation (collection) to their consumption. Through illustration, we gave a review of some published models that we compare based on their contribution to the creation of OGD's values namely politic, economic and social values. Finally, we conclude that there are still some data aspects that need to be included in the design of a successful data lifecycle including data interoperability and user feedback as key factors in the achievement of the objectives expected from OGD.

REFERENCES

[1] Morgan Currie (2018), A Dual Valuation of Open Government. In dg.o '18 Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, Article No. 75, Delft, The Netherlands — May 30 - June 01, 2018.

[2] Office of the Press Secretary The White House. 2013. Executive Order - Making Open and Machine Readable the New Default for Government Information.(May2013).https://www.whitehouse.gov/the-press office/2013/05/09/ executive-order-making-open-and machine-readable-new-default-government-

[3] Jos van Wezel, Achim Streit, A. Garcia, Andreas Heiss (2012), Data Life Cycle Labs, A New Concept to Support Data-Intensive Science, Published 2012 in ArXiv.

[4] Stuti SAXENA (2017), Nature and Scope of Open Government Data (OGD) in Oman, Journal of Social and Administrative Sciences Volume 4 Issue 2 June .

[5] Roberto D. Triviño (2016), State of Open Government Data as a Process in Ecuador, Conference: 2016 Third International Conference on eDemocracy & eGovernment (ICEDEG), At Sangolqui, Ecuador.

[6] Chiehyeon Lim, Ki-HunKim, Min-JunKim, Jun-YeonHeo, Kwang-JaeKim, Paul P.Maglio (2018), From data to value: A nine-factor framework for data-based value creation in information-intensive services, International Journal of Information ManagementVolume 39, April 2018, Pages 121-135.

[7] Claudio Sonaglio Albano (2013), Open government data: a value chain model proposal. In dg.o '13 Proceedings of the 14th Annual International Conference on Digital Government Research Pages 285-286, Quebec, Canada — June 17 - 20, 2013.

[8] Mohammad Alamgir Hossain, Yogesh K. Dwivedi & Nripendra P. Rana (2016), State-of-the-art in open data research: Insights from existing literature and a research agenda, Journal of Organizational Computing and Electronic Commerce Volume 26, 2016 - Issue 1-2: Open Data: Value Creation for Policy Makers and Analysts, Pages 14-40.

[9] Carmelo Francesco Origlia, Domenico Cersosimo Tito Bianchi, Alfredo Fortunato (2016), Assessing an open

government data initiative, The case of OpenCoesione. In Conference for E-Democracy and Open Government (CeDEM), Krems, Austria, 18-20 May 2016.

[10] Muhammad Anshari, Mohammad Nabil Almunawar, Syamimi Arif Lim (2018), Big Data and Open Government Data in Public Services. In ICMLC 2018 Proceedings of the 2018 10th International Conference on Machine Learning and Computing Pages 140-144, Macau, China — February 26 - 28, 2018.

[11] Adegboyega Ojo, Marijn Janssen (2014), Workshop about the understanding and improving the uptake and utilization open data. In dg.o '14 Proceedings of the 15th Annual International Conference on Digital Government Research Pages 350-351 Aguascalientes, Mexico — June 18 - 21, 2014.

[12] Florian Bauer, Martin Kaltenböck (2012), Linked Open Data: The Essentials A Quick Start Guide for Decision Makers, ed. mono/monochrome, January 2012.

[13]https://sunlightfoundation.com/policy/documents/ten-open-data-principles/ (Accessed 24/12/2018).

[14] Amir Sinaeepourfard, Jordi Garcia, Xavier Masip-Bruin, Eva Marín-Torder (2016), Towards a Comprehensive Data LifeCycle Model for Big Data Environments. In BDCAT '16 Proceedings of the 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies Pages 100-106, Shanghai, China — December 06 - 09, 2016.

[15] Judie Attard, Fabrizio Orlandi, Sören Auer (September 2016), Data Driven Governments: Creating Value Through Open Government Data, Transactions on Large-Scale Data- and Knowledge-Centered Systems XXVII: Special Issue on Big Data for Complex Urban Systems (pp.84-110).

[16] Judie Attard, Fabrizio Orlandi, Sören Auer (2016), Value Creation on Open Government Data, 2016 49th Hawaii International Conference on System Sciences, Koloa, HI, USA,5-8 Jan. 2016.

[17]https://www.w3.org/2011/gld/wiki/GLD_Life_cycle (Accessed 24/12/2018).

[18] Beth Plale, Inna Kouper (2017), Chapter 4 – The Centrality of Data: Data Lifecycle and Data Piplines, Data Analytics for Intelligent Transportation Systems 2017, Pages 91-111.

[19] Fatemeh Ahmadi Zeleti, Adegboyega Ojo, Edward Curry (2016), Exploring the economic value of open government data,Government Information Quarterly Volume 33, Issue 3, July 2016, Pages 535-551.

[20] Pin-Yu Chu, Hsien-Lee Tseng (2018), Open Data in Support of E-governance Evaluation: A Public Value Framework. In ICEGOV '18 Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance Pages 338-343 Galway, Ireland — April 04 - 06, 2018.

[21] Ricardo Matheus, Marijn Janssen, Devender Maheshwari (2018), Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities, Government Information Quarterly Available online 7 February 2018.

[22] Martin Lnenicka, Jitka Komarkova (2019), Developing a government enterprise architecture framework to support the requirements of big and open linked data with the use of cloud computing, International Journal of Information Management Volume 46, June 2019, Pages 124-141.

[23] Megan K. Sutherland, Meghan E. Cook (2017), Data-Driven Smart Cities: A Closer Look at Organizational, Technical and Data Complexities. In dg.o '17 Proceedings of the 18th Annual International Conference on Digital Government Research Pages 471-476, Staten Island, NY, USA — June 07 - 09, 2017.

[24] Ku, M. & Gil-Garcia, J. R. (2015). Challenges for Evidence-Based Decision Making to Combat Urban Blight: Revisiting Local Government Information Management through the Lens of Code Enforcement. Paper presented at the Association for Public Policy Analysis and Management.

[25] Judie Attard, Fabrizio Orlandi, Sören Auer (2016), Data Value Networks: Enabling a New Data Ecosystem, 2016 IEEE/WIC/ACM International Conference on Web Intelligence, Omaha, NE, USA, 13-16 Oct. 2016.

[26]Iryna Susha, Marijn Janssen, Stefaan Verhulst, Theresa Pardo (2017), Data Collaboratives: How to Create Value from Data for Public Problem Solving?. In dg.o '17 Proceedings of the 18th Annual International Conference on Digital Government Research Pages 604-606, Staten Island, NY, USA — June 07 - 09, 2017.

[27] Sharon S.Dawes, Lyudmila Vidiasova, Olga Parkhimovich (2016), Planning and designing open government data programs: An ecosystem approach, Government Information Quarterly Volume 33, Issue 1, January 2016, Pages 15-27.

[28] M. El Arass,I. Tikito,N. Souissi (2017), Data lifecycles analysis: towards intelligent cycle ,2017 Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 17-19 April 2017.

[29] A. Sinaeepourfard, X. Masip-Bruin, J. Garcia, and E Marín-Tordera (2015), "A Survey on Data Lifecycle Models: Discussions toward the 6Vs Challenges," Technical Report (UPC-DACRR- 2015-18), 2015