Journal of Business and Management Sciences, 2021, Vol. 9, No. 4, 181-191
Available online at http://pubs.sciepub.com/jbms/9/4/4
Published by Science and Education Publishing
DOI:10.12691/jbms-9-4-4

Exploring the Maturity of Open Governments in Various Countries: An Approach of Machine Learning

Po-Yuan Shih[1], Ting-Wei Wu[2,*], Cheng-Ping Cheng[1]

[1]Department of Finance, National Yunlin University of Science & Technology, Douliu, Yunlin 64002, Taiwan
[2]Department of Information Management, National Yunlin University of Science & Technology, Douliu, Yunlin 64002, Taiwan
*Corresponding author: wutingw@yuntech.edu.tw

Received October 17, 2021; Revised November 20, 2021; Accepted December 01, 2021

Abstract In recent years, open data movements have launched around the world. Open data have a wide range of participation and applications in daily activities, business areas, and government policies. Many countries believe that organizations or individuals accessing data on the data platform can develop new insights and innovations to enhance the lives of citizens. The concept of an open government (OG) is born with this context. Many countries publicize the public sector's data for citizens to download and create new applications. However, the government's open data platform may even lead to problems in lawsuits due to the inconsistency or damage in the data. Therefore, how to examine the maturity of an open government is a strong desire or demand of all countries. This study tries to define the maturity of an open government from an ICT (Information and Communication Technology) and open data development perspective. By collecting with different country's IDI (ICT Development Index) and GODI (Global Open Data Index) from 2015 to 2016, the maturity of an open government is classified into three categories with machine learning approach. Discussion on cluster members changed in national regions between 2015 and 2016, and suggestions of how to increase the maturity of OG or prevent a decline in maturity in countries are presented.

Keywords: *open government, open data, machine learning, maturity, Global Open Data Index (GODI), ICT Development Index (IDI)*

Cite This Article: Po-Yuan Shih, Ting-Wei Wu, and Cheng-Ping Cheng, “Exploring the Maturity of Open Governments in Various Countries: An Approach of Machine Learning.” *Journal of Business and Management Sciences*, vol. 9, no. 4 (2021): 181-191. doi: 10.12691/jbms-9-4-4.

1. Introduction

Open Data has gradually entered people’s lives in these years. Open Data means that data are freely accessible, usable, shared, modified and established by anyone for any purpose [1]. Open Data are different from Big Data, the value of Open Data lies in data sharing and can be used by other third parties, and respond to a series of specifications on technology, economics and law [2]. During these years, the volume and diversity of the open data released by public sector also increased [1].

Open Data Initiative was signed by President Obama in 2009 and Open Data Charter by G8 in 2013. Many government’s open data platform provide citizens and stakeholder with government information about the region or country involved [3]. The concept of the open government (OG) has gradually matured, the idea of OG by the Whitehouse, include publish government information online, improve the quality of government data, and establish a framework for OG that can be executed. The US Federal Government's Open data platform (Data.gov) was also launched in May 2009.

However, the maturity of the OG data platform will also affect the difficulty to use, like inconsistent data format, lack of metadata, there also no complete model to evaluate these open data platforms [4]. Users also encounter new entering difficulties without platform guidelines or data corruption and inconsistent while downloading data, with the complex data usage rights make it impossible for users to use these data arbitrarily, even encounter lawsuits problems seriously [5].

There are many aspects research on OG issues, like the meaning of the data itself there are different perspectives for different stakeholders, for comprehend the data through scientific, political and economic perspectives [6], or exploring the adoption situations of open data in different public sectors [7]. Additionally, the related research on the maturity of the OG platform, most of them have been developed diverse OG maturity models through different perspectives for evaluate the public sectors or measure the benchmarks of other similar platforms [8,9,10].

Table 1. OG assessment

OG Maturity	evaluation target	Topic	Author
Social media use	health care management organization	OG Maturity	Lee & Kwak [8]
	Arab government public sector		Schwalje & Aradi [9]
Open data perspective	Bangladesh OGD adoption	OD Assessment	Talukder et al. [24]
	US Open data Platform	OD Assessment	Veljković et al. [10]
Multivariate model	Mexico OG Website	OG Assessment	Sandoval-Almazan & Gil-Garcia [21]
OG innovation	Government Innovation Literature Review	Open Innovation	Ham et al. [22]
Transparency and Accountability perspective	Reform of Australia OG	OG Assessment	Henninger [25]
	OG data platforms	OD Assessment	Lourenço [23]
Systematic review	OG data platforms	OD Assessment	Attard et al. [3]
Case study	Korea's e-government	OG Maturity	Sangki [26]

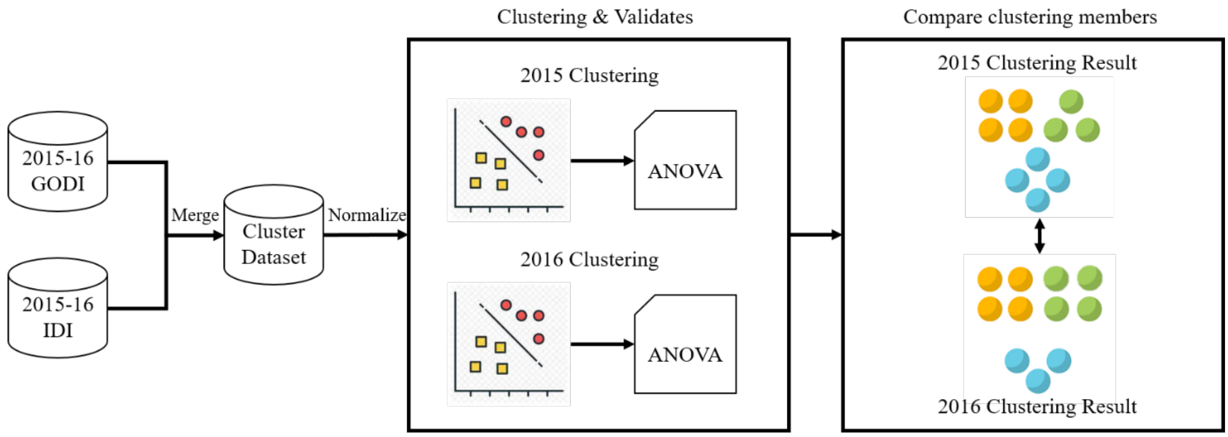

Figure 1. Overall clustering analysis scenario

When discussing the overall market economy, indicator research, or level and stage of development, similar individuals are often clustered in groups. For example, World Bank Atlas method which is based on GNI per capita, Four groups were divided into Low-income economies to High-income economies. Human Development Index (HDI) issued by Human Development Reports is calculated based on the different ethnic income, education level, and average life expectancy of each country, from Very high human development to Low human development, it was divided into 4 groups. Melin et al., [11] Use Self Organizing Maps to divide the severity of coronavirus in major cities around the world into 4 groups (Very High, High, Medium, and Low).

The Global Open Data Index of OG established for the UK's non-profit organization Open Knowledge Foundation (OKFN), this index is based on open data for each country, Therefore, the scores of the evaluations are also highly related to the maturity of the OG. In addition, the investment and development of ICT and the productivity of the country have a positive relationship with the economy. Many countries have invested a lot of resources to develop.

ICT. In response to the digital age, OG services such as data platforms, websites and other services have a relation of links with ICT development. Therefore, this study defines the maturity of the OG as the index performance of IDI and GODI, and exploring the maturity of OGs in various countries. By adapt general index for OG data such as the Global Open Data Index (GODI) and the ICT Development Index (IDI) from ITU (International Telecommunication Union), this study tries to use machine learning technique in clustering three categories in the maturity of OG development, and provides suggestions in countries' strategy they may take.

2. Background and Preliminary

2.1. Conceptual of Open Data

According to ODI (Open Data Institute) definition of open data, means data can be accessed by anyone, and must have a license, if the license missing, the data will not be reused. The detailed description of the license is that when the user employs the data, it must indicate who published. In addition, when users mix open data with other dataset and publish, they must also publish as open data. Hounsell et al. [12] use open data from Public Transportation Information to create a friendly interface mobile application to help older travelers, and Fermoso et al. [13] link cultural heritage knowledge with local government open data and enrich the sources of information of local sectors are all related research of open data.

2.2. Open Government Definition

The concept of an OG can be traced back to the 1950s, according to the research by Yu & Robinson [14], an issue "The Open Government Principle: Applying the People's Right to Know" is the first to mention the terms about the OG. However, it may come from the memorandum on transparent and OG signed by US President Obama in 2009. It mentions that the government must be transparent, involved and collaborative [15]. More detailed information such as publishing government information on the Internet, improving the quality of government information, establishing and institutionalizing an OG culture and establishing an OG framework that can be implemented [16]. Obama signed an executive order to produce open and machine-readable government data in 2013, which included a description of the open data policy [17]. Open data are the heart of the government movement, allowing other organizations or individuals to access data to develop new insights and innovations and enhance the lives of others, freely available government data can be used to create useful tools and products and help people live more easily in modern life. Veljković et al. [10] proposed a benchmark evaluation of the US federal government's open data platform (Data.gov) with an open data perspective or understand the different meanings of OG data from the perspective of different stakeholders [6]. Whitmore [18] uses past contract information to predict war and explore the inconvenience for the users of the immature open data platform.

2.3. Definition of OG Maturity

For the maturity of OG, we can first describe the e-government maturity study. According to Harrison et al. [19], the OG is an extension of the e-government's future public management situation, and Fath-Allah et al. [20] compiled 25 models of e-government maturity and proposed three classifications. It’s also mentioned that most OG maturity studies are used to evaluate and rank for the OG platforms.

Related research on the maturity and evaluation of the OG, Lee & Kwak [8] proposes a public participation OG maturity model based on social media, the evaluation target is the health care management organization in the United States. Veljković et al. [10] benchmarking the OG platform with an open source perspective. Sandoval-Almazan & Gil-Garcia [21] evaluated the 32 OG websites in Mexico using the multicomponent model. Ham et al. [22] explored how the government proposes an open innovation maturity model based on an open systems perspective through the context of people's participation and innovation of open data. Lourenço [23] uses transparency and accountability perspective to analyze OG data platforms in many countries. The overall list of OG maturity and assessment studies is as follows (Table 1).

3. OG Maturity in Progress

3.1. ICT Development Index

According to Niebel [27], the investment in Information and Communication Technologies (ICT) is seen as a key driver of productivity growth. In the past few years, many countries have also tried to develop ICT through a large amount of resources, more affluent countries are also considered to have more resources on ICT, and reached higher level [28]. In addition, other countries have implemented ICT-related policies and invested in a large number of ICT infrastructures in schools [29]. And, ICT itself also covers a wide range of technologies, to assess ICT development at different stages. Many studies tend to use absolute scores [30,31,32]. IDI is a comprehensive index and contains 11 sub-indicators as an evaluation standard to monitor or compare the development of ICT between different countries. This index was developed in 2008 by the International Telecommunication Union (ITU) and was released in 2009 as the first edition. The main purpose of IDI is to assess the long-term ICT development level and assessment in each country, digital gap in ICT maturity, and can use them to evolve and develop in the existing situation.

3.2. Global Open Data Index

The Global Open Data Index (GODI) is an annual measure of OG data for each country. This index was sought by volunteers from the Open Knowledge International organization, and the initial release date is October 28, 2013. It’s from the perspective of citizens to observe and record. In the government open data collections observed, the definitions of the datasets observed each year are slightly different. The survey team defines that each open dataset must have key characteristics such as geographic information, how often the data is updated, etc. The score ranges are from 5 to 30 points for each item, with a total score of 100 points (as in Table 2).

Table 2. GODI Government open data set definition

Data Set	Description
National Statistics	Population or economic indicators. It is necessary to include regular updates on national GDP, population, and unemployment statistics.
Government Budget	Government budget and expenditure, government and public sector needs to updated budget regularly.
Government Spending	Regularly provide detailed transaction records and past government spending
Legislation	Relevant content such as national laws, plaintexts and norms can be accessed online
Election Results	Relevant results for voters and voting in each major constituency
National Map	Need to have high-quality national map data, including proportions, national boundaries, and other data
Pollutant Emissions	Information need for air pollution damage to human health, for example, PM, carbon monoxide (CO), etc.
Company Register	Regularly provide a list of company registrations
Location dataset	Data set of zip code and area code, such as Zipcodes Addresses and Coordinates data.
Government procurement tenders	National or state government collects all relevant information such as tender and awards
Water Quality	Water source and quality and disease prevention
Weather forecasty	It is required to provide information such as temperature, wind speed and precipitation every 5 days and annual historical temperature.
Land Ownership	Cadastral data showing land ownership must include land boundaries, dimensions, and owner’s name
Transport Timetables	Public transport schedule, updated annually
Health Performance	The location and opening hours of public hospitals and medical facilities, infectious disease rate and updated at least once a year

Source: Global Open Data Index (http://index.okfn.org/methodology/).

4. Methodology and Analysis Result

4.1. Clustering Technique of Machine Learning

Grouping different objects for different purpose can be seen in different areas, whether it is engineering, science, medical, anthropology, etc. In real case, data usually without labels in advance, most of it are grouping together by calculating the similarity, which is called clustering in machine learning [33]. About the definition of clustering, Xu & Tian [34] believes that members of the same cluster are like each other, and not similar in different clusters. Saxena et al., [33] expresses the clustering architecture through a subset of S_1, S_2, ..., S_k in a set S as shown as follow:

$$S_1 \cap S_2 \cap S_3, \ldots, \cap S_k = \varnothing \quad (1)$$

This means that any member in S (S_1, …, S_k) just happens to have only one sub collection and does not belong to other sub collections. Clustering also means that people can identify the data through some basic similarities through the key characteristics. In the technical method of clustering, it can be divided into two types: hierarchical and partitional method, the hierarchical algorithm works by presenting a nested pattern and a similarity level group through a dendrogram. The partitional clustering algorithm is to obtain a single partition of the data as a clustering structure. The advantage of the partition method is that in a larger amount of data, the presentation is better than the hierarchical algorithms [35].

4.1.1. Hierarchical Clustering (HC) Methods

The hierarchical clustering method can be divided into two forms, agglomerative and divisive hierarchical clustering, agglomerative clustering follows a bottom-up approach where data or clusters are merged sequentially from the bottom of the tree structure. The divisive hierarchical clustering method follows a top-down approach, splitting clusters containing all objects into smaller clusters, the hierarchical grouping method is mainly as follows.

4.1.1.1. Single-linkage Clustering

This type of algorithm is also known as connectedness, minimum method, or nearest neighbor method, means that a link between pairs of elements in a cluster that is the closest. In this method, the distance method between the two clusters is identified by the nearest distance method, and the distance between the members of any cluster and other cluster members is the closest. This also defines the similarity. If the data has similarity, the similarity of the paired members in a cluster will be higher than the paired members between the other clusters. If the 2 clusters A and B are presented as shown in the equation below [33]:

$$\mathrm{Min}\{d(a,b) : a \in A, b \in B\} \quad (2)$$

4.1.1.2. Complete-linkage Clustering

Complete-linkage clustering is also known as the maximum method or the furthest neighbor method. Compared to Single-linkage clustering, in Complete-linkage clustering, members of either cluster of two clusters take the longest distance from members of another cluster, If the 2 clusters A and B are presented as shown in the equation below [33]:

$$\mathrm{Max}\{d(a,b) : a \in A, b \in B\} \quad (3)$$

4.1.2. Partition Clustering Methods

Partition clustering moves members from one cluster to another. These methods typically require the user to set up the cluster. To obtain global optimality in a partition-based clustering approach, a detailed enumeration process is required for all possible partitions [36].

4.1.2.1. K-means Clustering

K-means is one of the most well-known and simple algorithms at present, and it is also the most commonly used method for clustering, in this method, the data requires the user to set the cluster k in advance. The main concept of the method is to define the center point of k for each cluster. The main function J is as follows:

$$\text{Minimize J} = \sum_{j=1}^{k}\sum_{i=1}^{n} X_i^{(j)} - C_j^{\ 2} \quad (4)$$

where $X_i^{(j)} - C_j^{\ 2}$ is a chosen distance measure between a data point $X_i^{(j)}$ and the cluster center C_j [33].

4.1.2.2. EM Clustering

EM is a quite mature clustering algorithm in the statistical community, it is also an unsupervised clustering method of machine learning commonly used for data point density estimation. Steps of Expectation (E) and Maximization (M) are performed iteratively till the results converge. E step computes an expectation of the likelihood by including the latent variables, and M step computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found on the last E step [37].

Table 3. Variables in this study

Input Variable			
Variable	**Variable name**	**Data source**	**Annotation**
x_1:IA	ICT access	IDI (ITU)	
x_2:IU	ICT use	IDI (ITU)	
x_3:IS	ICT skills	IDI (ITU)	
x_4:NS	National Statistics	GODI	
x_5:PT	Procurement tenders	GODI	
x_6:NM	National Map	GODI	
x_7:DL	Draft Legislation	GODI	
x_8:PE	Pollutant Emissions	GODI	Merge to AQ in 2016
x_9:ER	Election Results	GODI	
x_{10}:CR	Company Register	GODI	
x_{11}:GB	Government Budget	GODI	
x_{12}:WF	Weather forecast	GODI	
x_{13}:WQ	Water Quality	GODI	
x_{14}:GS	Government Spending	GODI	
x_{15}:LD	Location datasets	GODI	
x_{16}:LO	Land Ownership	GODI	New variable in 2016
x_{17}:NL	National Laws	GODI	New variable in 2016
x_{18}:AB	Administrative Boundaries	GODI	New variable in 2016
x_{19}:AQ	Air Quality	GODI	New variable in 2016

4.2. Data Acquisition and Processing

According to the report published by the ITU, Open Data Index official website presents information from 2013 to 2016. However, due to the lack of information in 2014, and the number of countries include by the Open Data Index in 2013 is small, this study can only use the IDI and GODI index from 2015 to 2016. Therefore, there are 15 variables in 2015 and 18 variables in 2016, as shown in Table 3.

4.3. Clustering of OG Maturity and Verification from ANOVA Analysis

Figure 1 is the overall clustering analysis scenario in this study. According to Madhulatha [38], clustering can be said to be the most important unsupervised learning problem in machine learning. Common clustering methods for partitioned algorithms are K-mean, EM (expectation maximization), and so on.

According to research by Jung, Kang & Heo [39], the k-mean and EM clustering methods are used to cluster the experimental data. Therefore, we use WEKA 3.8 algorithm to cluster IDI and GODI index of each country from 2015 to 2016. Another research from Ayanso & Lertwachara [31] is adopted to set the maturity of OG into 3 groups base on their center score of cluster which are Matured OG, OG in transition and Maturing OG. The cluster result of distribution between 2015 and 2016 are shown in Table 4 and Table 5 below, while Figure 2 and Figure 3 are the OG cluster boxplot of the two-year period. Table 6 is a detailed group file for each year. The values in Table 6 are the averages center of 3 categories calculated by the EM grouping algorithm. It is easy to classify that the members of the Matured OG have a fairly high score in the IDI and GODI index in comparison with member of OG in transition and Maturing OG cluster.

Table 4. 2015 OG maturity cluster distribution

2015 OG Distribution		
Matured OG (29%)	**OG in transition (32%)**	**Maturing OG (39%)**
1. Australia	1. Albania	1. Antigua and Barbuda
2. Austria	2. Argentina	2. Barbados
3. Brazil	3. Belgium	3. Bolivia
4. Canada	4. Bulgaria	4. Botswana
5. Colombia	5. Chile	5. Cambodia
6. Czech Republic	6. Greece	6. Cameroon
7. Denmark	7. Hong Kong	7. Costa Rica
8. Finland	8. India	8. Dominican Republic
9. France	9. Indonesia	9. El Salvador
10. Germany	10. Israel	10. Guatemala
11. Italy	11. Jamaica	11. Guyana
12. Mexico	12. Japan	12. Iran
13. Netherlands	13. Latvia	13. Kenya
14. Norway	14. Portugal	14. Macedonia
15. Romania	15. Russian	15. Malaysia
16. Sweden	16. Singapore	16. Myanmar
17. United Kingdom	17. Slovakia	17. Nepal
18. United States	18. Switzerland	18. Oman
19. Uruguay	19. Thailand	19. Pakistan
	20. Turkey	20. Panama
	21. Ukraine	21. Paraguay
		22. Philippines
		23. South Africa
		24. Tanzania
		25. Trinidad and Tobago
		26. Tunisia

Table 5. 2015 OG maturity cluster distribution

2016 OG Distribution		
Matured OG (41%)	**OG in transition (24%)**	**Maturing OG (35%)**
1. Argentina	1. Albania	1. Antigua and Barbuda
2. Australia	2. Bolivia	2. Barbados
3. Austria	3. Bulgaria	3. Botswana
4. Belgium	4. El Salvador	4. Cambodia
5. Brazil	5. Greece	5. Cameroon
6. Canada	6. India	6. Costa Rica
7. Chile	7. Panama	7. Dominican Republic
8. Colombia	8. Paraguay	8. Guatemala
9. Czech Republic	9. Portugal	9. Guyana
10. Denmark	10. Romania	10. Indonesia
11. Finland	11. Russia	11. Iran
12. France	12. South Africa	12. Jamaica
13. Germany	13. Switzerland	13. Kenya
14. Hong Kong	14. Thailand	Macedonia
15. Israel	15. Turkey	Malaysia
16. Italy	16. Ukraine	Myanmar
17. Japan		Nepal
18. Latvia		Oman
19. Mexico		Pakistan
20. Netherlands		Philippines
21. Norway		Tanzania
22. Singapore		Trinidad and Tobago
23. Slovakia		Tunisia
24. Sweden		
25. United Kingdom		
26. United States		
27. Uruguay		

Table 6. 2015 OG clustering result

2015 Matured OG variables					
IA	**IU**	**IS**	**NS**	**PT**	**NM**
0.78	0.75	0.81	0.99	0.73	0.8
DL	**PE**	**ER**	**CR**	**GB**	**WF**
0.73	0.56	0.67	0.45	0.92	0.54
WQ	**GS**	**LD**	**LO**		
0.34	0.17	0.64	0.37		
2016 Matured OG variables					
IA	**IU**	**IS**	**NS**	**PT**	**NM**
0.79	0.76	0.82	0.93	0.66	0.67
DL	**ER**	**CR**	**GB**	**WF**	**WQ**
0.65	0.61	0.55	0.95	0.66	0.3
GS	**LD**	**LO**	**NL**	**AB**	**AQ**
0.18	0.41	0.11	0.74	0.75	0.74
2015 OG in transition					
IA	**IU**	**IS**	**NS**	**PT**	**NM**
0.64	0.55	0.75	0.63	0.5	0.44
DL	**PE**	**ER**	**CR**	**GB**	**WF**
0.52	0.53	0.51	0.34	0.6	0.43
WQ	**GS**	**LD**	**LO**		
0.28	0.19	0.2	0.24		
2016 OG in transition					
IA	**IU**	**IS**	**NS**	**PT**	**NM**
0.5	0.44	0.66	0.73	0.67	0.35
DL	**ER**	**CR**	**GB**	**WF**	**WQ**
0.48	0.25	0.38	0.73	0.3	0.12
GS	**LD**	**LO**	**NL**	**AB**	**AQ**
0.16	0	0.11	0.57	0.6	0.35
2015 Maturing OG					
IA	**IU**	**IS**	**NS**	**PT**	**NM**
0.35	0.28	0.43	0.38	0.39	0.19
DL	**PE**	**ER**	**CR**	**GB**	**WF**
0.38	0.12	0.36	0.08	0.45	0.2
WQ	**GS**	**LD**	**LO**		
0.06	0.07	0.12	0.07		
2016 Maturing OG					
IA	**IU**	**IS**	**NS**	**PT**	**NM**
0.36	0.32	0.4	0.55	0.34	0.1
DL	**ER**	**CR**	**GB**	**WF**	**WQ**
0.22	0.1	0.06	0.56	0.07	0
GS	**LD**	**LO**	**NL**	**AB**	**AQ**
0	0.03	0.01	0.32	0.13	0.14

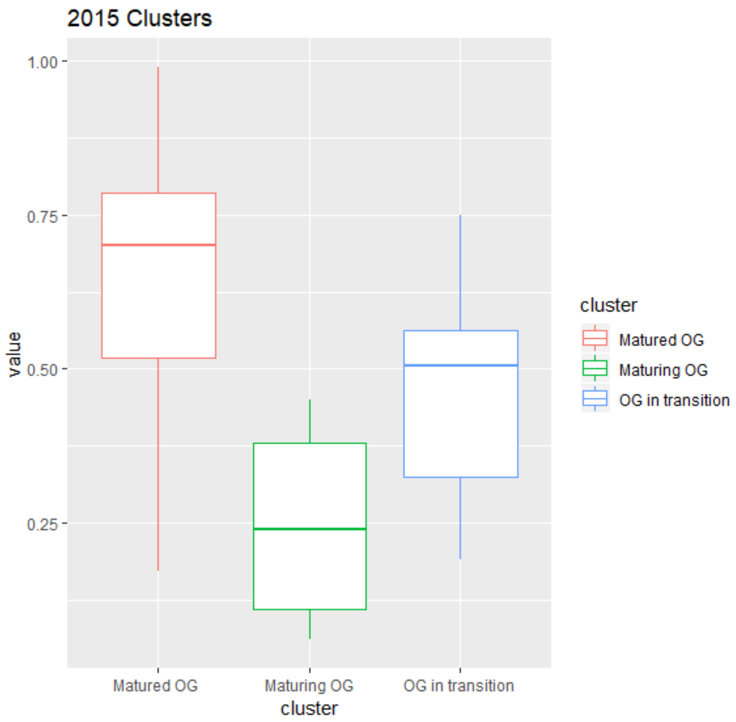

Figure 2. 2015 OG cluster boxplot visualization

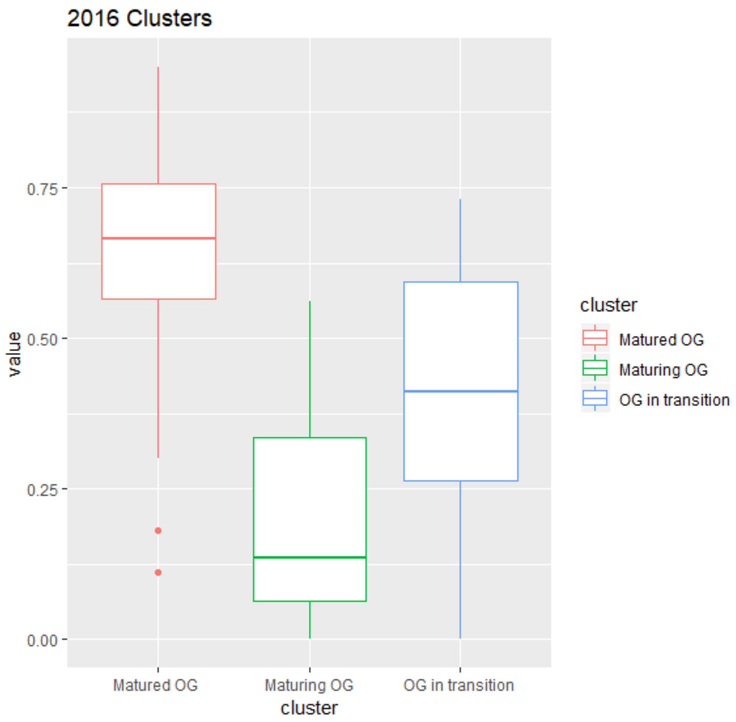

Figure 3. 2016 OG cluster boxplot visualization

In order to verify the effect of clustering, ANOVA analysis with the cluster model as a covariate was performed to verify the significance [40,41]. Repeated-measures analyses of variance with one factor (marker cluster) and three levels (Maturing OG, OG in transition and matured OG) were used to compare the clustering result [42]. The probability of type I error was set as $\alpha = 0.05$, there was significantly difference between the three clusters through the ANOVA analysis no matter in Table 7 or Table 8. And, that indicated that our proposed method in clustering result is acceptable.

4.4. Discussion on Cluster Member's Changes in National Regions

After separated cluster analysis of 2015 and 2016 in OG maturity, the members of the cluster may change their OG maturity level due to unexpected reason. Therefore, we compared the clustering results between two years and found 4 kinds of situation changed as in Table 9. Whereas (↑) means an upgrade of OG maturity and (↑) is a downgrade of OG maturity in Table 9. This study attempts to discuss the situation changed in different clusters.

Table 7. ANOVA analysis of 2015 cluster

2015 Cluster	Cluster Mean (SD)			ANOVA		Contrasts		
Variables	**Maturing OG(C1)**	**OG in transition(C2)**	**Matured OG(C3)**	**P**	**F**	**C1*C2**	**C1*C3**	**C2*C3**
NS	0.379 (0.265)	0.631 (0.118)	0.99 (0.315)	<0.001*	62.715	<0.001*	<0.001*	<0.001*
PT	0.387 (0.283)	0.505 (0.184)	0.734 (0.294)	<0.001*	9.951	0.273	<0.001*	0.019
NM	0.19 (0.209)	0.441 (0.203)	0.797 (0.267)	<0.001*	39.812	0.001*	<0.001*	<0.001*
DL	0.381 (0.229)	0.524 (0.141)	0.724 (0.224)	<0.001*	16.013	0.056	<0.001*	<0.05*
PE	0.121 (0.189)	0.536 (0.222)	0.563 (0.228)	<0.001*	32.315	<0.001*	<0.001*	0.912
ER	0.362 (0.264)	0.51 (0.29)	0.668 (0.436)	<0.001*	4.773	0.284	0.08	0.288
CR	0.085 (0.145)	0.343 (0.285)	0.45 (0.268)	<0.001*	10.836	<0.05*	<0.001*	0.436
GB	0.45 (0.268)	0.591 (0.239)	0.916 (0.146)	<0.001*	23.038	0.101	<0.001*	<0.001*
WF	0.2 (0.229)	0.426 (0.254)	0.542 (0.362)	<0.001*	8.759	0.021	<0.001*	0.398
WQ	0.064 (0.111)	0.281 (0.267)	0.345 (0.254)	<0.001*	10.989	<0.05*	<0.001*	<0.001*
GS	0.071 (0.043)	0.191 (0.249)	0.168 (0.206)	<0.05*	2.943	0.070	0.182	0.921
LD	0.119 (0.189)	0.2 (0.32)	0.637 (0.376)	<0.001*	18.453	0.622	<0.001*	<0.001*
LO	0.075 (0.125)	0.241 (0.21)	0.366 (0.274)	0.111	11.454	0.020	<0.001*	0.136
IA	0.349 (0.231)	0.638 (0.249)	0.777 (0.18)	<0.001*	21.759	<0.001*	<0.001*	0.131
IU	0.283 (0.198)	0.554 (0.237)	0.75 (0.191)	<0.001*	28.140	<0.001*	<0.001*	0.012
IS	0.435 (0.209)	0.752 (0.159)	0.806 (0.114)	<0.001*	32.421	<0.001*	<0.001*	0.585

Table 8. ANOVA analysis of 2016 cluster

2016 Cluster	Cluster Mean (SD)			ANOVA		Contrasts		
Variables	**Maturing OG(C1)**	**OG in transition(C2)**	**Matured OG(C3)**	**P**	**F**	**C1*C2**	**C1*C3**	**C2*C3**
GB	0.583 (0.275)	0.73 (0.273)	0.948 (0.12)	<0.001*	16.906	0.133	<0.001*	<0.05*
NS	0.561 (0.26)	0.719 (0.152)	0.943 (0.116)	<0.001*	26.525	<0.05*	<0.001*	0.001*
PT	0.354 (0.235)	0.675 (0.183)	0.663 (0.343)	<0.001*	9.695	<0.05*	<0.05*	0.989
NL	0.313 (0.224)	0.625 (0.219)	0.73 (0.2)	<0.001*	24.686	<0.001*	<0.001*	0.273
AB	0.163 (0.276)	0.603 (0.359)	0.739 (0.378)	<0.001*	18.561	<0.05*	<0.001*	0.422
DL	0.226 (0.245)	0.488 (0.321)	0.646 (0.246)	<0.001*	15.656	<0.05*	<0.001*	0.148
AQ	0.148 (0.294)	0.375 (0.37)	0.735 (0.192)	<0.001*	28.082	<0.05*	<0.001*	<0.001*
NM	0.122 (0.202)	0.347 (0.352)	0.676 (0.318)	<0.001*	22.710	0.054	<0.001*	<0.05*
WF	0.087 (0.197)	0.306 (0.254)	0.669 (0.24)	<0.001*	40.873	0.013	<0.001*	<0.001*
CR	0.05 (0.086)	0.388 (0.408)	0.567 (0.389)	<0.001*	15.894	<0.05	<0.001	0.193
ER	0.096 (0.255)	0.306 (0.425	0.598 (0.443)	<0.001*	10.842	0.217	<0.001*	<0.05*
LD	0.261 (0.086)	0 (0)	0.422 (0.442)	<0.001*	15.895	0.958	<0.001*	0.276
WQ	0 (0)	0.131 (0.246)	0.307 (0.36)	<0.001*	8.750	0.276	<0.001*	0.089
GS	0 (0)	0.166 (0.358)	0.187 (0.334)	<0.05*	3.159	0.166	0.053	0.967
LO	0.130 (0.063)	0.109 (0.208)	0.113 (0.222)	0.111	2.279	0.231	0.128	0.998
IA	0.356 (0.243)	0.521 (0.229)	0.796 (0.159)	<0.001*	28.385	<0.05*	<0.001*	<0.001*
IU	0.325 (0.201)	0.456 (0.215)	0.766 (0.151	<0.001*	36.882	0.085	<0.001*	<0.001*
IS	0.409 (0.22)	0.673 (0.207)	0.826 (0.126)	<0.001*	32.460	<0.001*	<0.001*	<0.05*

Table 9. Cluster change scenario

Group	Country	Cluster-2015	Cluster-2016	2015 GODI Rank	2016 GODI Rank	2015 IDI Rank	2016 IDI Rank
Situation A (↑)	Bolivia	Maturing OG	OG in transition	69	53	107	111
	El Salvador			71	49	106	118
	Panama			88	61	89	93
	Paraguay			50	37	112	109
	South Africa			54	43	88	88
Situation B (↑)	Argentina	OG in transition	Matured OG	54	17	52	55
	Belgium			35	22	21	22
	Chile			29	22	55	56
	Hong Kong			37	24	9	6
	Israel			44	41	36	30
	Japan			31	16	11	10
	Latvia			31	14	37	40
	Slovakia			23	17	19	20
	Singapore			50	32	47	42
Situation C (↓)	Indonesia	OG in transition	Maturing OG	41	61	108	115
	Jamaica			37	58	105	99
Situation D (↓)	Romania	Matured OG	OG in transition	13	24	59	60

4.4.1. Situation A

The Open Government Partnership (OGP) was established in 2011, the eight founding countries are Brazil, Indonesia, Mexico, Norway, the Philippines, South Africa, the United Kingdom, and the United States.

South Africa in Table 9, the rank of GODI in 2015 was 54 and it rose to 43 in 2016. The sub-indicator ICT use in the IDI index score rose from 0.31 to 0.42. ICT use is related to the use of Internet, fixed and mobile bandwidth by individual users. According to CIPESA (The Collaboration on International ICT Policy in East and Southern Africa). OG not just opening up the data of various departments, ICT is one of the key drivers of the OG. It also mentions the large increase in the population of the African people using modern communication technologies such as the internet and mobile phones. The interaction between people through ICT has also increased.

Cernuzzi & Pane [43] pointed out that Paraguay has a population of about 6.8 million, which is the region with the lowest GDP growth in Latin America. Until 2012, only 27% of users used the Internet and 29% of households owned a computer. But in recent years, the government of Paraguay has placed IT on the policy agenda with the goal of improving public management processes, providing quality public services, through IT-based solutions, Paraguay also became a member of OGP in 2012. Paraguay government continue to provide IT infrastructure and build tools to help people find and use open data and move toward the goal of OG [44]. It is known that the Paraguayan government continues to work hard to open data and OG goals.

Panama lost slightly to Paraguay in overall GODI performance. It has also increased from 88 to 61 in GODI. According to the Panama 2015-2016 progress report provided by the Independent Reporting Mechanism (IRM), Panama officially participated in the OGP activities in 2012. As members of the OGP, Panama officially filed and drafted 20 commitment proposals in April 2015. It was found that the Panamanian government had a slight deficiency in the standardized digital format and proposed to cultivate and strengthen the use and value of open data for citizens and government agencies. Therefore, Panama still has a way to go in the development of OG data.

4.4.2. Situation B

Slovakia signed into OGP in 2011, hope that better transparency can prevent corruption and promote citizen participation, and such a move also has a positive effect, although the process of establishing an open data platform was quite difficult. Slovakia released more than 600 data sets from 26 government departments in 2015. Slovakia's IT services, the National Agency for Network and Electronic Services (NASES) and the Office of the Plenipotentiary for the Development of Civil Society all claim that Slovakia's open data have a bright future [45].

Latvia in GODI's performance, ranking 31 in 2015 and rising to 14th in 2016. The European Data Portal's website also constantly mentions that Latvia has continued to open new data platforms and combine political efforts to increase the availability of open data. The attention of the Argentine government in the development of open data has increased from 54 in 2015 to 17th in 2016 in GODI ranking. The OECD pointed out that Argentina’s public spending is the highest in Latin America, accounting for 44% of GDP, this growth is mainly caused by recurring expenditures. And, Argentina’s financial pressure is increasing, the budget deficit in 2015 also accounted for 7.4% of GDP. However, the total number of public employments in Argentina is 18%, which is the third largest employed population in the region. The introduction of ICT in the procurement system has made great progress and established an e-sourcing platform and announced where there are procurement opportunities, but how to fully profit from it and deal with lower ICT knowledge and innovation is a big challenge.

The OECD Country Fact Sheet in Belgium shows that Belgian citizens are very satisfied with the local health care and education system, and the satisfaction is higher than the average of the OECD group surveys. It is also mentioned that Belgium established a new Federal Public

Service in 2017 to organize IT, human resources, budget and other items. With the development of the Internet of Things in recent years. The application of smart cities has followed, Zotano & Bersini [46] based on the ontology of smart cities in Belgium and established a data-driven methodology for assessing the potential of smart cities based on different open data platforms. It can be seen that the combination of the mature development of ICT and open data may be one of the development goals of smart cities in the future.

As a member of the OGP, Chile's report on Chile Open Government Action Plan 2016-2018 proposes to promote the quality improvement and integration of citizens in energy management, seeking cooperation and participation in justice on environmental issues, civic education programs, taking into account open data policies and directives, strengthening open data policies and encouraging citizens to access information in the public sector, and strengthen and promote the commitment of public procurement tools.

Israel joined ODP in 2012 and the first national action plan was implemented in 2012-2013, and Israel's 2016 GODI's National Spending and National Statistics variables also show full scores. The overall ranking has also increased from 44 in 2015 to 41 in 2016, but the legal access and other evaluation items are still possible to make progress.

Japan's open data platform data.go.jp was established in 2013, and the OG data strategy of the Japan Open Data Charter Action Plan in the platform mentioned four principles. It includes the government need to disclose the data online, disclose data in a machine-readable format, to encourage the use of the data for commercial and non-commercial use, to have specific measures for the disclosure of the data, and the results must be steadily cumulative. In Japan, the performance of GODI in 2016 was much better than that in 2015. The four variables were full marks and the ranking rose from 31 to 16.

4.4.3. Situation C

Although Indonesia is one of the founding members of OGP, its performance in the cluster is declining, from OG in transition to Maturing OG. According to Gunawan & Amalia [47], study on the implementation of open data in Indonesia indicates that although most regions are willing to disclose data, but most of the information is in PDF format, and even Microsoft Word and Microsoft PowerPoint formats, and users also need to spend more effort to get the data. The study also pointed out that no area releases RDF data, and if want to open the RDF format need relevant knowledge, and some of these open data sets are difficult to understand since there are lack of evaluation in the legal access, computer-readable data format, and data download ability. It can be seen that Indonesia needs to improve in the development of the OG.

4.4.4. Situation D

Jamaica joined OGP in 2016 and opened its first open data platform in June. Up to now, no national action plan has been proposed, and there is no relevant commitment to promote OG. It may also lead to the stagnation of Jamaica's progress in OG.

Romania joined OGP in 2011 and has also implemented two action plans. The current 2016-2018 action plan is also underway, but the 2016 GODI rankings are down compared to 2015. The lack of evaluation in machine-readable formats accounted for the majority, the data needs to be registered by the user before downloading and partial link damage may result in a decline in ranking.

5. Analyses and Suggestions

5.1. Developing ICTs and Action Plans can Improve the OG Maturity

According to the clustering results, the scores of GODI and IDI indicators of the members in the Matured OG cluster are both high. To climb from the OG in transition cluster to the Matured OG, a certain extent of ICT development is necessary. Restakis et al. [48] argue that ICTs contribute to government governance and civil rights, and that ICT's architecture and transmission services enhance and protect citizens' and human rights and further transparency and encourage citizens to participate in broader democratic practices. Fumega [49] mentioned that most OG data organizations were established after 2005, and they have technical components in their daily routines and plans. ICT is also considered a key enabler for a new generation of communication. The government's message to the people through social media is also a new form. Karakiza [50] proposes an OG maturity model that combines web 2.0 and social media tools in five levels, however the government's use of social media services is still in its early stages. Most governments focus on participation and transparency, but less on collaboration. Khan, Swar & Lee [51] also categorizes the use of social media to increase communication and citizen participation, transparency, etc.

In addition, joining an OG-related organization or developing a multi-year OG strategy can help improve the maturity of the OG. Ubaldi [52] puts forward many challenges in the integration of Open Government Data (OGD). Policy needs to be able to be intelligently disclosed and to addressed. Technically, the data formats and standards of government departments are not unified, and it's hard to making it impossible to access data efficiently from the perspective of users. The economic and financial challenges are key factors such as the need to convert data into a reusable format and store it on government websites in an open source manner. If every country follows open definitions on open data or use other tools to improve the format and standards of their data, it will also increase the maturity of OG.

5.2. Developing ICTs and Action Plans can Improve the OG Maturity

As in the cluster change situation in Table 9, cluster performance in Latin America shows signs of rising, such as Panama, Chile, Argentina, Paraguay and so on. Latin America is the highest proportion of countries participating in OGP. It also mentions the 94 OGP commitments for accessing public information in Latin America, and strengthens the existing laws for accessing information, can be seen in Latin America's efforts to OG. Scrollini & Ochoa [53] describes the state of Latin America when it comes to members of OGP who are involved in a National Action Plan (NAPs) to discuss with civil society groups. It also mentioned that the quality of NAPs in many regions needs to be improved, and the future promises of many NAPs may have been implemented or easily achieved, the government must be more ambitious in the commitment of NAPs. It also mentions that the possibility of formulating and implementing reforms depends on the situation of each country.

From the rankings in Table 9, it can be seen that the IDI index scores of the Situation A members who have been upgraded from Maturing OG to OG in transition have not improved significantly. However, there has been considerable progress in the performance of GODI, so improving the performance of GODI is the key to jumping off the Maturing OG cluster. In addition, if it is a member of the OG in transition cluster, it can be known that the IDI score of the cluster member is higher than that of the members of the Maturing OG cluster. ICT development is necessary because ICT can open up new ways of communication and citizen participation, as well as transparency and civil rights [48].

6. Conclusions

This study collects GODI and IDI indicators of multiple countries from 2015 to 2016 for OG maturity investigation. By using machine learning technique, the data was clustered into three groups named Maturing OG, OG in transition and Matured OG. The results of this study show that the GODI and IDI index scores are relatively high in the Matured OG cluster, and the OG maturity of the Latin American region has improved significantly. This study can give advice to decision makers through an overall data analysis, which are ICT development and investment, joining OG-related organizations, and developing OG action plans can help improve the maturity of the government. However, members of different clusters have different promotion strategy priorities. Members who are themselves in the Maturing OG cluster should first focus on improving GODI scores, working with social citizen groups and developing relevant action plans can help improve GODI scores. In addition to developing action plans, members of the OG in transition cluster should invest in the ICT development and innovative services which will help to enhance the maturity of the OG. Therefore, policy makers can formulate their follow-up OG development strategies in the future based on the maturity of OG in their countries.

The latest GODI information is only released until 2016, therefore, this study is the only research that exploring OG maturity base on IDI and GODI index until now. In the future, other indicators may be included for further consideration, Such as ODIN (The Open Data Inventory) Score, IPB (ICT Price Basket) or Open Data Barometer, etc. OGP's declaration of OG also includes support for citizen participation, but the part of national citizen participation is often difficult to quantify its performance. Therefore, it is more feasible to evaluate the maturity of each country in the OG from the content quality of OG data and the performance of ICT development. It is hoped that the follow-up decision makers will be provided with improvements for future OG development.

References

[1] Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. Government Information Quarterly, 33(2), 325-337.

[2] Chignard, S. (2013). A brief history of open data. Paris Tech Review, 29.

[3] Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. Government Information Quarterly, 32(4), 399-418.

[4] Huijboom, N., & Van den Broek, T. (2011). Open data: an international comparison of strategies. European journal of ePractice, 12(1), 4-16.

[5] Beno, M., Figl, K., Umbrich, J., & Polleres, A. (2017). Open data hopes and fears: Determining the barriers of open data. In E-Democracy and Open Government (CeDEM), 2017 Conference for (pp. 69-81). IEEE.

[6] Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. Government Information Quarterly, 32(4), 441-452.

[7] Wang, H. J., & Lo, J. (2016). Adoption of open government data among government agencies. Government Information Quarterly, 33(1), 80-88.

[8] Lee, G., & Kwak, Y. H. (2012). An open government maturity model for social media-based public engagement. Government Information Quarterly, 29(4), 492-503.

[9] Schwalji, W., & Aradi, W. (2013). An Arab open government maturity model for social media engagement. Tahseen Consulting. Accessed on, 1-19.

[10] Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. Government Information Quarterly, 31(2), 278-290.

[11] Melin, P., Monica, J. C., Sanchez, D., & Castillo, O. (2020). Analysis of Spatial Spread Relationships of Coronavirus (COVID-19) Pandemic in the World using Self Organizing Maps. Chaos, Solitons & Fractals, 109917.

[12] Hounsell, N., Shrestha, B., McDonald, M., & Wong, A. (2016). Open data and the needs of older people for public transport information. Transportation Research Procedia.

[13] Fermoso, A. M., Mateos, M., Beato, M. E., & Berjón, R. (2015). Open linked data and mobile devices as e-tourism tools. A practical approach to collaborative e-learning. Computers in Human Behavior, 51, 618-626.

[14] Yu, H., & Robinson, D. (2012). The New Ambiguity of "Open Government". Princeton CITP. Yale ISP Working Paper.

[15] Obama, B. (2009). Transparency and Open Government, memorandum for the heads of executive departments and agencies. Retrieved from Web site (http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment).

[16] Orszag, P. (2009). Open government directive. Memorandum for the Heads of Executive Departments and Agencies, 2.

[17] Obama, B. (2013). Executive Order--Making Open and Machine Readable the New Default for Government Information. The White House.

[18] Whitmore, A. (2014). Using open government data to predict war: A case study of data and systems challenges. Government Information Quarterly, 31(4), 622-630.

[19] Harrison, T. M., Guerrero, S., Burke, G. B., Cook, M., Cresswell, A., Helbig, N., ... & Pardo, T. (2012). Open government and e-government: Democratic challenges from a public value perspective. Information Polity, 17(2), 83-97.

[20] Fath-Allah, A., Cheikhi, L., Al-Qutaish, R. E., & Idri, A. (2014). E-government maturity models: A comparative study. International Journal of Software Engineering & Applications, 5(3), 71.

[21] Sandoval-Almazan, R., & Gil-Garcia, J. R. (2014). Towards an evaluation model for open government: A preliminary proposal. In International Conference on Electronic Government (pp. 47-58).

[22] Ham, J., Lee, J. N., Kim, D., & Choi, B. (2015). Open innovation maturity model for the government: an open system perspective, Thirty Sixth International Conference on Information Systems, Fort Worth 2015

[23] Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. Government Information Quarterly, 32(3), 323-332

[24] Talukder, M. S., Shen, L., Talukder, M. F. H., & Bao, Y. (2019). Determinants of user acceptance and use of open government data (OGD): An empirical investigation in Bangladesh. Technology in Society, 56, 147-156.

[25] Henninger, M. (2018). Reforms to counter a culture of secrecy: Open government in Australia. Government Information Quarterly, 35(3), 398-407.

[26] Sangki, J. (2018). Vision of future e-government via new e-government maturity model: Based on Korea's e-government practices. Telecommunications Policy, 42(10), 860-871.

[27] Niebel, T. (2018). ICT and economic growth–Comparing developing, emerging, and developed countries. World Development, 104, 197-211.

[28] Doong, S. H., & Ho, S. C. (2012). The impact of ICT development on the global digital divide. Electronic Commerce Research and Applications, 11(5), 518-533.

[29] De Witte, K., & Rogge, N. (2014). Does ICT matter for effectiveness and efficiency in mathematics education? Computers & Education, 75, 173-184.

[30] Dragomirescu, H., & Sharma, R. S. (2009). Operationalizing the sustainable knowledge society concept through a multi-dimensional scorecard. In World Summit on Knowledge Society (pp. 328-337).

[31] Ayanso, A., & Lertwachara, K. (2015). An analytics approach to exploring the link between ICT development and affordability. Government Information Quarterly, 32(4), 389-398.

[32] Nath, H. K., & Liu, L. (2017). Information and communications technology (ICT) and services trade. Information Economics and Policy, 41, 81-87.

[33] Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., ... & Lin, C. T. (2017). A review of clustering techniques and developments. Neurocomputing, 267, 664-681.

[34] Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. Annals of Data Science, 2(2), 165-193.

[35] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. ACM computing surveys (CSUR), 31(3), 264-323.

[36] Rokach, L., & Maimon, O. (2005). Clustering methods. In Data mining and knowledge discovery handbook (pp. 321-352). Springer, Boston, MA.

[37] Naik, D., & Shah, P. (2014). A review on image segmentation clustering algorithms. Int J Comput Sci Inform Technol, 5(3), 3289-93.

[38] Madhulatha, T. S. (2012). An overview on clustering methods. IOSR Journal of Engineering, Vol. 2(4) pp: 719-725

[39] Jung, Y. G., Kang, M. S., & Heo, J. (2014). Clustering performance comparison using K-means and expectation maximization algorithms. Biotechnology & Biotechnological Equipment, 28(sup1), S44-S48.

[40] Calinski, T., & Corsten, L. C. A. (1985). Clustering means in ANOVA by simultaneous testing. Biometrics, 39-48.

[41] Harb, H., Makhoul, A., & Couturier, R. (2015). An enhanced K-means and ANOVA-based clustering approach for similarity aggregation in underwater wireless sensor networks. IEEE Sensors Journal, 15(10), 5483-5493.

[42] Magalhães, F. A., Souza, T. R., Araújo, V. L., Oliveira, L. M., de Paula Silveira, L., de Melo Ocarino, J., & Fonseca, S. T. (2020). Comparison of the rigidity and forefoot–Rearfoot kinematics from three forefoot tracking marker clusters during walking and weight-bearing foot pronation-supination. Journal of Biomechanics, 98, 109381.

[43] Cernuzzi, L., & Pane, J. (2014). Toward open government in Paraguay. IT Professional, 16(5), 62-64.

[44] Steinberg, M., & Castro, D. (2017). The State of Open Data Portals in Latin America. Retrieved from Web site (https://www.datainnovation.org/2017/07/the-state-of-open-data-portals-in-latin-america/).

[45] Jordan Schneider (2015), Bringing Government Data into the Light: Slovakia's Open Data Initiative,2011-2015. Retrieved from Website:https://successfulsocieties.princeton.edu/sites/successfulsocieties/files/JS_OGP_Slovakia_FORMATTED_012Oct2015.pdf.

[46] Zotano, M. A. G., & Bersini, H. (2017). A data-driven approach to assess the potential of Smart Cities: the case of open data for Brussels Capital Region. Energy Procedia, 111, 750-758.

[47] Gunawan, D., & Amalia, A. (2016). The Implementation of open data in Indonesia. In Data and Software Engineering (ICoDSE), 2016 International Conference on (pp. 1-6). IEEE.

[48] Restakis, J., Araya, D., Calderon, M. J., & Murray, R. (2014). ICT, open government and civil society. Journal of Peer Production, 7.

[49] Fumega, S. (2017). Rethinking the Crucial Impact of ICT in Civil Society Organisations Working in FOI and OGD Fields. In Proceedings of the 18th Annual International Conference on Digital Government Research (pp. 485-494). ACM.

[50] Karakiza, M. (2015). The impact of social media in the public sector. Procedia-Social and Behavioral Sciences, 175, 384-392.

[51] Khan, G. F., Swar, B., & Lee, S. K. (2014). Social media risks and benefits: A public sector perspective. Social Science Computer Review, 32(5), 606-627.

[52] Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives. OECD Working Papers on Public Governance, (22).

[53] Scrollini, F., & Durand Ochoa, U. (2015). Perspectives on open government in Latin America. Strategic Update (15.1). LSE IDEAS, London School of Economics and Political Science, London, UK.