Hindawi
Complexity
Volume 2022, Article ID 5876035, 14 pages
https://doi.org/10.1155/2022/5876035

Research Article

A Comprehensive Method for Improving the Quality of Open Government Data and Increasing Citizens' Willingness to Use Data by Analyzing the Complex System of Citizens and Organizations

Mohammad Moradi,[1] Mojtaba Mazoochi,[1] and Mohammad Ahmadi [2]

[1]*ICT Research Institute, Tehran, Iran*
[2]*Faculty of Engineering and Computer Science, Department of Software engineering, Khatam Al-Nabieen University, Kabul, Afghanistan*

Correspondence should be addressed to Mohammad Ahmadi; mohammad.moradi@ut.ac.ir

Received 5 May 2022; Revised 3 June 2022; Accepted 3 August 2022; Published 21 August 2022

Academic Editor: Andrea Murari

In recent years, the amount of data in the world is growing rapidly. Data growth also occurs in the government sector. All ministries and institutions at every level are data producers. These government-owned data have a high potential if they can be used properly. Open government data can stimulate innovation and economic growth and enhance business models. In order to increase the willingness of citizens to use open government data and enjoy the benefits mentioned, the quality of open government data needs to be improved. The quality of open government data encompasses a variety of dimensions and criteria. Also, the importance of each dimension and criterion in increasing the quality of open government data is different. Therefore, we are faced with a complex system that requires proper decision-making and management. In fact, we are dealing with decision-making in the complex management system. Given the importance of this issue, the purpose of this study is to provide a new and comprehensive method to improve the quality of open government data and increase the willingness of citizens to use the data by considering the complex network of citizens and organizations. For this purpose, library studies have been used to extract comprehensive and effective dimensions and criteria. The statistical population includes all articles related to the criteria of improving the quality of open government data and increasing the willingness of citizens to use the data. The probabilistic sampling method of simple random samples has been used, and 10 articles in this field have been reviewed. After extracting the criteria as well as the data of 112 governmental organizations and institutions related to each criterion from the open data portal, the complex network of citizens and governmental organizations and institutions has been analyzed in order to identify high-degree centrality organizations. Then, the data characteristics of the organizations that were most desired by the citizens were extracted using data mining techniques including the regression model. Also, field method and multicriteria decision-making technique including the DEMATEL technique have been used to express the solutions and identify the cause-and-effect relationships between the solutions. The criteria extracted in improving the quality of open government data and increasing the willingness of citizens to use the data are included: "data originality," "license openness," "up-to-datedness," "data access," "metadata completeness," "number of data sets," "format openness," "nondiscrimination," "understandable," "number of categories of data sets," "free," "lack of missing data," "data request ability," "visualization," "feedback," and "data subject matter." Based on the results obtained from the analysis of the complex network and the regression model, the criterion of "society subject" with a coefficient of 72.564 and a positive sign has the greatest impact on increasing the number of citizens' visits to open government data. After that, the criterion of "format openness" with a coefficient of 52.682 and a positive sign has the second rank in increasing the number of visits. Extracting comprehensive and effective criteria in improving the quality of open government data and increasing citizens' willingness to use data, calculating the weight and importance of each criterion by analyzing the complex network of citizens and organizations, as well as providing solutions, can help managers in decision-making and proper management in the complex system of citizens and government organizations.

1. Introduction

Open data refer to nonconfidential data that are made available without any restrictions on use or distribution [1]. Open government data are tools for empowering citizens and giving them access and permission to use data generated by the government sector, so that they can use, store, redistribute, and integrate data with other data sources [2]. In addition, open government data can be defined as data belonging to a government entity that is published for free use, reuse, and redistribute [3].

Providing information in the form of open data will reduce corruption, gain public trust, and build a democratic society. Open data provide more opportunities to monitor governance activities. For example, it makes the way the budget is spent transparently and the effects clear. It also encourages citizens to be more involved in overseeing governance. In the corporate sphere, open data primarily help the entity itself to be aware of the existence of data in the organization and to avoid parallel and costly activities to collect data that have already been done by the organization. Open government data enable citizens to participate in decision-making processes with informed and structured procedures [4]. Open government data can stimulate innovation and economic growth and enhance business models [5].

In order to increase citizens' desire for open government data and enjoy the benefits mentioned, the quality of open government data needs to be improved. The quality of government data encompasses a variety of dimensions and criteria. Also, the importance of each dimension and criterion in increasing the quality of open government data is different. Therefore, we are faced with a complex system that requires proper decision-making and management. In fact, we are dealing with decision-making in complex management systems. In this study, first, comprehensive and effective dimensions and criteria in increasing the quality of open government data and increasing the willingness of citizens to use the data have been extracted using library studies. In the next phase, an attempt has been made to determine the extent of citizens' willingness to use the data of each organization by presenting a new and creative method and considering the complex network of citizens and government organizations and institutions providing open data. Then, the characteristics of the data that the citizens wanted were examined, and based on the obtained results, the importance and weight of the dimensions and effective criteria in increasing the quality of open government data were calculated. For this purpose, data mining techniques and regression model have been used. Also, after analyzing the results, multicriteria decision-making technique has been used to provide solutions and their importance. The method presented in this study and the results obtained can help managers in the decision-making and proper management of organizations providing open government data to increase data quality and thus increase the desire of citizens to use and enjoy the benefits of open data (as a kind of decision-making in the complex system of citizens and organizations providing open government data).

In the following, in Section 2, the related works and the disadvantages and limitations of previous researches and the reason for dealing with the current research are stated. Section 3 describes the research method, which consists of different phases and includes the analysis of a complex network of citizens and organizations to calculate the weight and importance of criteria for improving the quality of open government data and increasing citizens' willingness to use the data. In Section 4, the results are presented based on the different phases expressed in the research methodology section and the findings are discussed. Findings include the extraction of comprehensive dimensions and criteria in improving data quality and increasing citizens' willingness to use data, extraction of data of government organizations and institutions based on each criterion, results of complex network analysis of citizens and organizations and identification of organizations with high centrality, use of data mining techniques involves a regression model to analyze the characteristics of the data set related to organizations with high degree centrality and calculate the weight and importance of each characteristic and criterion, and to provide solutions and identify the cause-and-effect relationships between the solutions using decision-making trial and evaluation laboratory (DEMATEL) technique. Finally, in Section 5, the conclusion is given.

2. Related Works

In this section, research conducted to improve the quality of open government data and increase the willingness of citizens to use the data is reviewed. Nikiforova and McBride conducted a study entitled "Open government data portal usability: A user-centered usability analysis of 41 open government data portals." Confirming the importance of portal usability for the data reuse process, this study helps to explain some of the initial insights by asking two questions: "How can the usability of open government data portals be evaluated and compared in different contexts?" and "What are usually the practical aspects of open government data portals?". To answer these research questions, a set of 41 open government data portals have been selected for usability analysis based on the feedback of 40 users. According to the results of this study, the lack of interaction between users with open government data portals in cases such as providing feedback or requesting data sets is one of the main problems of open government data portals. Therefore, governments should focus on developing open government data ecosystems and increasing the interoperability of these portals [6].

Zhang and Xiao have examined the framework for evaluating the quality of open government data. The purpose of this study is to create a common framework as a reference for evaluating the quality of open government data. In this research, 10 qualitative studies have been combined in a

common reference framework to evaluate the quality of open government data. Based on a seven-step analysis, a common reference framework for evaluating the quality of open government data is presented, which includes six criteria: accuracy, accessibility, completeness, up-to-datedness, stability, and comprehensibility [7].

De Juana-Espinosa and Luján-Mora evaluated open government data portals in the EU from 2015 to 2017. This study presents data collected from open government data portals in 28 EU countries. Several parameters and criteria observed over a period of 3 years in open national data portals have been identified and recorded to create this data set. The data are obtained manually from existing public information sources and official open government data portals that are freely available on the Web. In this study, the criteria "Existence of a link from an open government data portal to the source site of the data set provider," "Existence of social network plugins" to discuss users' experiences in using an open government data portal, "Support for various data set formats" and "Data search and filtering capability" have been proposed as criteria for evaluating open government data portals [8].

Zheng et al. in a study entitled "Evaluating global open government data: Methods and status" first examines seven methods of evaluating open government data by regularly comparing and analyzing their frameworks, criteria, and methods. Based on this analysis, a framework for evaluating the performance of open government data has been developed for all UN member states. According to the results, most of the current evaluation programs focus on data and foundation and pay less attention to software platforms, use and impact. This study shows that in 2018, 34 countries (18%) score "very high," 40 countries (21%) score "high," 43 countries (22%) score "medium," while 76 countries (39%) have received a "low" score [9].

Dahbi et al. have conducted a study entitled "Toward an evaluation model for open government data portals." In this study, the authors define an evaluation model for open government data portals based on several main dimensions that have a great impact on their application. Specified dimensions are information richness that deals with adapting the portal to the needs of the user in terms of content; detection capability associated with tools and mechanisms that increase data access to the portal; reusability, which deals with the openness of the data published on the portal and the possibility of reusing them; and interaction, which is related to the openness of the portal for user feedback, cooperation, and interaction with published data. The proposed evaluation model has been used to evaluate four national open government data portals [10].

Vetrò et al. offer an approach to measuring the quality of open government data sets. They asked open government data users about the challenges to open government data quality. In general, they suggest that the data set be evaluated for completeness, accuracy, traceability, comprehensibility, compliance, and expiration. In other words, the ideal data set should include complete and accurate data, be machine-readable, have metadata, be updated, be accurate, and be traceable in terms of source. This is a very complete and useful framework for policymakers who want to control the quality of data sets in open government projects or programs [11].

Dawes et al. introduced a framework for evaluating the quality of open data portals at the national level and provided a set of criteria for evaluating data quality problems in open government data portals. These criteria were applied to 12 portals, and several dimensions of data quality were introduced. These dimensions included the existence of standards in data formats, the existence of metadata, machine readability, and the up-to-datedness of data [12].

Misuraca and Viscusi discuss the framework for evaluating the compliance of open government data based on quality. These criteria include three dimensions of quality: completeness, accuracy, and up-to-datedness [13]. Harrison et al. focus on evaluating metadata quality [14].

As seen in previous research, each of the studies focused on a specific dimension of improving the quality of open government data and increasing citizens' willingness to use the data. Also, in the researches done so far, the weight and importance of each of the dimensions and criteria have not been specified. Considering the weight and importance of each dimension and criterion based on the degree of willingness of citizens to use the data so far has not been considered in previous research. In this study, different and comprehensive dimensions and criteria in improving the quality of open government data and increasing the willingness of citizens to use data as well as calculating the importance of each criterion based on the analysis of a complex network of citizens and open data providers are presented. Also, after reviewing the obtained results, the solutions and the causal relationships between them are discussed.

3. Materials and Methods

In this research, the type of research based on the purpose is applied research. In the first phase, library studies are conducted to extract comprehensive dimensions and criteria for improving the quality of open government data and increasing citizens' willingness to use the data. The statistical population includes all articles related to the dimensions and criteria for improving the quality of open government data and increasing the willingness of citizens to use the data. The probabilistic sampling method of simple random samples has been used, and 10 articles in this field have been reviewed.

In the second phase, the organizations, including all government organizations and institutions present in the open data portal (https://data.gov.ir/), are examined and the data related to each criterion extracted in the first phase are calculated for each organization. The number of government organizations and institutions surveyed was 112.

In the third phase, the complex network of citizens and government organizations and institutions providing open data is analyzed. This network is a directional network. The nodes in this network are citizens and government organizations and institutions, and the links represent data visits. For example, in Figure 1, the citizen nodes are shown in

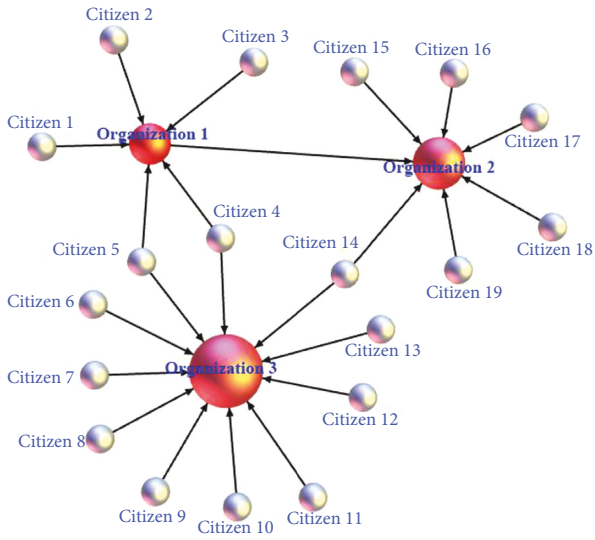


Figure 1: A complex network of citizens and government organizations and institutions.

silver and the organizations nodes are shown in red. The link from citizen i to organization j indicates that citizen i has visited the published open data of organization j. In this network, there can be directional links from citizen to an organization or organization to another organization. But there is no possibility of directional link between citizens or organization to citizen in this network. It is also possible to create a link from one citizen to several organizations, meaning that one citizen can view data from multiple organizations.

Degree centralization has been used to identify organizations whose data are more desirable. The absolute degree centrality of node vi is calculated as follows [15]:

$$c_D(i) = \text{degree of vertex } i. \tag{1}$$

The relative degree centrality of node vi can be calculated as follows [15]:

$$C_D(i) = \frac{c_D(i)}{(n-1)}. \tag{2}$$

In formula (2), $n-1$ is the largest possible degree of a network with n nodes. This information can be extracted through the open data portal based on the number of visits to the data set that each organization has provided openly.

In the fourth phase, data mining techniques were used to extract the data characteristics of the organizations that were more interested (had a higher degree centrality). For this purpose, regression model has been used. The criteria extracted in the first phase were used as attributes, and the data extracted in the second phase, which are the data of organizations based on each criterion, were used as a record. The "visit rate" attribute has been used as a label. The output of this model is the coefficients related to each quality criterion that determine the weight and importance of that criterion.

In the fifth phase, the results are reviewed and solutions are presented. For this purpose, field methods and multicriteria decision-making techniques have been used. The DEMATEL technique was used to identify the cause-and-effect relationships between the solutions. The research process diagram is shown in Figure 2.

Also, the reasons for using each method stated in the research process are summarized in Table 1.

4. Results and Discussion

This section deals with the results of the five phases described in the previous section and discusses them.

4.1. Extracting Comprehensive and Effective Criteria in the Quality of Open Government Data and Increasing the Willingness of Citizens to Use the Data. In this section, comprehensive and effective criteria for the quality of open government data and increasing the willingness of citizens to use data in three dimensions of open data, data transparency, and interaction are stated. References related to each dimension and criterion are also specified.

4.1.1. Open Data

(1) Data Accuracy [*10*, *16*]. This dimension deals with the originality of the data, the absence of lost data, and the up-to-datedness. Missing data prevent the use and reuse of data and have a major impact on the quality of programs that reuse data. The following criteria relate to the data accuracy dimension:

(a) Data originality [11, 16].

(b) Lack of missing data [10].

(c) Up-to-datedness [10, 16, 17]: this criterion evaluates the up-to-datedness of the data in the published data set. For each O_i organization, this score is calculated based on data published in the last five years using formula (3). T_j represents the number of data published in j years ago.

$$O_i = \sum_{j=0}^{4} \left(1 - \frac{j}{10}\right) \times T_j. \tag{3}$$

(2) Discoverability [*10*]. Discoverability dimension deals with tools and mechanisms that increase data access and search. In other words, users should be able to search and access the relevant data set in a simple and efficient way. This will not happen if the metadata is not provided. Metadata provides a better understanding of the importance of data and data structure and helps users access the data they need [18]. Assessing the discoverability dimension requires a thorough evaluation of the descriptive metadata and the availability of data access features. The following criteria relate to the discoverability dimension.

(a) Metadata completeness [9, 10, 17, 19]: this criterion evaluates the completeness of descriptive metadata. For each data set, the completeness of the descriptive metadata fields is evaluated. These fields include title, description, tags, publisher, and more.

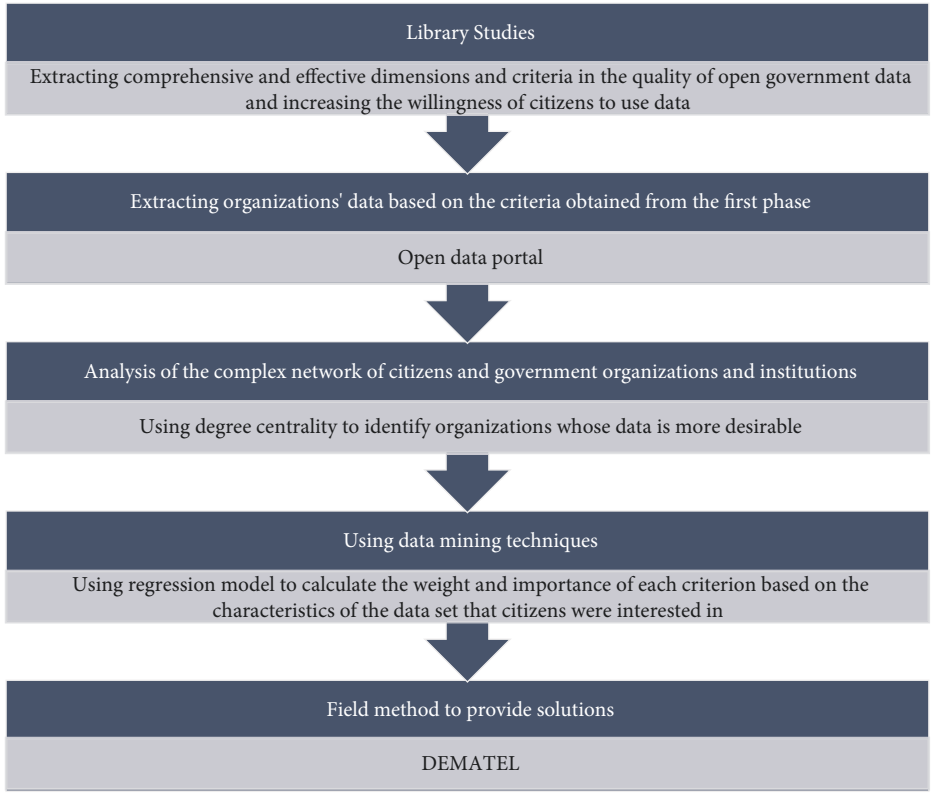


Figure 2: Diagram of the research process.

Table 1: Reasons for using each method in the research process.

Phase	Method	Reason for using the method
1	Library studies	The reason for using library studies has been to extract comprehensive and effective dimensions and criteria in the quality of open government data.
2	Extracting organizations' data based on the criteria obtained from phase 1	The reason for extracting the data of the organizations in the open data portal was to identify the characteristics of each organization. In fact, the data of this phase constitute attributes in phase 4.
3	Analysis of the complex network of citizens and organizations	In addition to the data extracted in phase 2, which identifies the characteristics of each organization, it is necessary to identify the organizations whose data have been most desired and used by citizens. For this purpose, the indegree of each organization in the complex network of citizens and organizations, which indicates the number of visits to the data of the organization, has been used. In fact, the data of this phase form the labels in phase 4.
4	Using data mining techniques	After extracting the data from phase 2 as attributes and the data from phase 3 as labels, it is necessary to analyze the data in order to identify the positive or negative impact of each attribute on the label (number of visits) and also to calculate the weight and importance of each attribute.
5	Field method to provide solutions	The reason for using this phase is to extract solutions and identify cause-and-effect relationships between the solutions in order to improve citizens' interaction with open government data. Identifying the cause-and-effect relationships between solutions will help managers spend more time and cost on the solutions that have the greatest impact on other solutions.

(b) Data access [9–11, 17]: this criterion evaluates the existence of attributes that increase data discovery, in particular, the existence of three attributes: search, sort, and filter, which receive their value in the range [0, 1] according to the existence of these attributes.

(3) Richness of Information [10]. Richness of information measures the extent to which a user needs are met in terms of the amount of data. The following criteria are related to the richness of information dimension.

(a) Number of data sets [10, 11, 17, 19]: this criterion evaluates the number of data sets that an organization has openly provided.

(b) Number of categories of data sets [19].

(c) Data subject matter [19]: citizens may be more interested in some issues. Based on the open data portal survey, the subjects of the entire data set were extracted as follows:

(i) Heights
(ii) Planning
(iii) Water effects
(iv) Animal and plant ecology
(v) Borders
(vi) Images/maps/land cover
(vii) Location
(viii) .Weather
(ix) Society
(x) Health
(xi) Management
(xii) Environment
(xiii) Farming
(xiv) Science and research education
(xv) Energy
(xvi) Structure
(xvii) Economy
(xviii) Transportation
(xix) Earth sciences

(d) Data request ability [7, 10]: this criterion measures the degree of openness to user requests. In other words, it examines the possibility of allowing users to request new data sets. Depending on the availability of the data request, two values of 1 or 0 are assigned.

4.1.2. Data Transparency

(1) Reusability [*10*, *16*]. The value of open government data is realized only after its reuse [3]. Open government data are considered reusable when the data are released under an open license and there is unrestricted access, reuse, and redistribution of data. It must also be published electronically and machine-readable. Reusability also deals with features that provide an easy way to reuse data, such as applications and the API (Application Programming Interface). The following criteria relate to the reusability dimension:

(a) License openness [10, 16]: this criterion evaluates the openness of the data set license for reuse.

(b) Format openness [10, 16, 19]: this criterion evaluates the openness of the data format. For each D_n data set, the FOI_n score based on the data format is assigned as follows:

- If the format is machine-unreadable: $FOI_n = 0$ (e.g., PDF).
- If the format is machine-readable: $FOI_n = 1$ (e.g., JSON, CSV).

(c) Free [16].

(d) Nondiscriminatory [11, 16]: access to and reuse of data are the same for all individuals and legal entities.

(2) Understandable [16]

4.1.3. Interactivity

(1) Feedback [*10*, *16*, *17*]. This criterion examines the existence of features related to collaboration, feedback, and evaluation and assesses the existence of three possibilities: commenting on the data set, ranking the data set, and feedback on the portal.

(2) Visualization [*10*, *17*, *19*]. This criterion evaluates the existence of visualization tools and features such as maps, diagrams, or programs for visualizing and interacting with data.

4.1.4. Chart of Comprehensive and Effective Criteria in the Quality of Open Government Data and Increase the Willingness of Citizens to Use the Data. Figure 3 shows a chart of comprehensive and effective criteria expressed in the quality of open government data and the increasing willingness of citizens to use the data based on the dimension, criteria, reference, and year of publication of the reference.

4.2. Extracting Organizations' Data Based on the Extracted Criteria. In this phase, the data of government organizations and institutions present in the open government data portal are extracted based on the criteria extracted in the previous phase. The number of organizations in this portal that provide open data was 112. Figure 4 shows a data extraction of government organizations and institutions present in the open government data portal based on the stated criteria.

4.3. Analysis of the Complex Network of Citizens and Government Organizations. In this phase, the complex network of citizens and government organizations and institutions present in the open government data portal were examined. In order to identify organizations and government institutions with a high degree centrality, input links from citizens to organizations that represent the indegree of each organization were calculated. This information was available based on the number of visits to each organization's data set in the open government data portal. Figure 5 shows a view of the extracted data, the number of visits to each organization, and its data set.

4.4. Using Data Mining Techniques to Calculate the Weight and Importance of Each Criterion. After extracting comprehensive and effective criteria in the quality of open government data and increasing the willingness of citizens to use the data, as well as extracting data from government organizations based on these criteria and calculating the number of visits to each organization's data set based on the complex network analysis of citizens and organizations, the weight and importance of each criterion need to be determined. As mentioned earlier in this study, we intend to

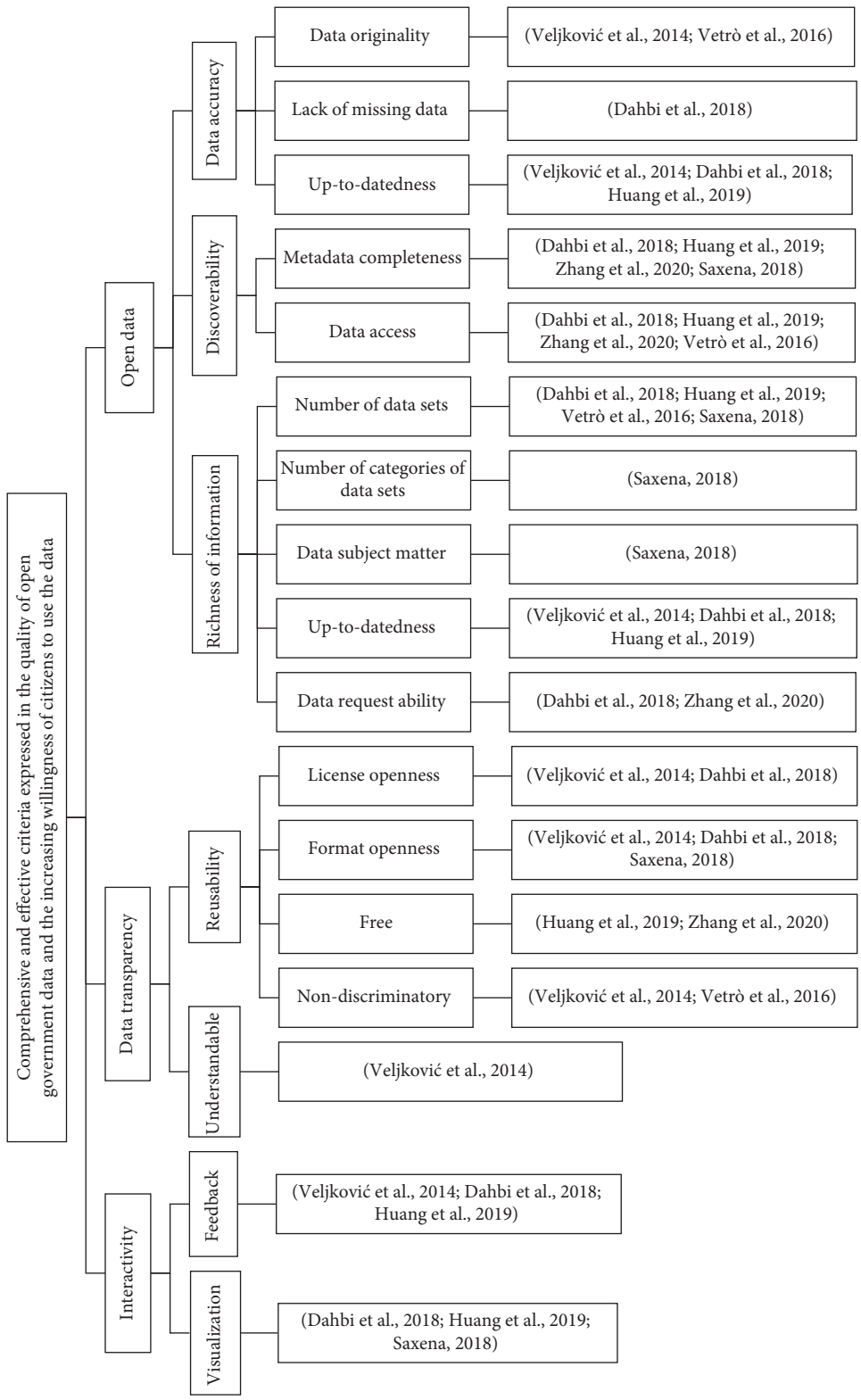


Figure 3: Chart of effective criteria in the quality of open government data and increasing the willingness of citizens to use data based on dimension, criteria, reference, and year of reference.

calculate the weight and importance of each criterion based on the degree of willingness that citizens have so far from the data set of each organization. In this way, first the data sets of the organizations that have been desired by the citizens are identified. Then, the characteristics of this data set are extracted, and based on this, the weight and importance of each criterion are determined. Therefore, the criteria extracted in the first phase are considered as attributes. The

I	H	G	F	E	D	C	B	A
)ata request ability	Up-to-datedness	Visualization	Format openness	Up-to-datedness	Data access	Number of categories of data sets	Number of data sets	Criterion name / Organization name
0	1765.6	1762	0	1765.6	223	3	2098	Geology organization
0	516	0	645	516	645	8	645	Statistical Center
0	235.9	36	226	235.9	226	1	337	Roads and Transportation Organization
0	73.5	0	0	73.5	0	5	81	Iran Health Insurance Organization
0	82.7	43	53	82.7	54	11	97	mmunication Technology Organization of
0	103.6	0	16	103.6	16	1	107	The official newspaper of the country
0	27.3	39	0	27.3	0	2	39	Lake Urmia Rehabilitation Headquarters
0	61	70	0	61	0	2	70	stitute of Communication and Information
0	46.8	52	0	46.8	0	6	52	ousing and Urban Development Research
0	28.3	46	0	28.3	0	10	46	Isfahan Regional Water Company
0	15.1	6	4	15.1	4	3	16	Social Security Organization
0	28.8	31	0	28.8	0	2	32	Railway Research Center
0	30.9	3	0	30.9	0	1	32	/ice President for Science and Technolog
0	18.9	20	0	18.9	0	1	20	National mapping agency
0	20.3	29	0	20.3	0	5	29	Space Agency
0	13.5	0	15	13.5	15	1	15	zation of Municipalities and Villages of the
0	9	0	0	9	0	1	10	Literacy Movement Organization
0	12.6	0	0	12.6	0	1	14	Airports and Air Navigation Company
0	13	0	0	13	0	1	13	Corporate Audit
0	22.8	0	0	22.8	1	3	24	Nomadic Affairs Organization
0	11.5	0	0	11.5	0	1	12	Space Research Institute

Figure 4: View of the extracted data of organizations based on each criterion.

H	G	F	E	D	C	B	A
ng data	Data originality	Data access	Metadata completeness	Data request ability	Number of categories of data sets	Number of data sets	Number of visits	Criterion name / Organization name
	2098	223	2098	0	3	2098	141118	Geology organization
	645	645	645	0	8	645	10355	Statistical Center
	337	226	337	0	1	337	6171	Roads and Transportation Organization
	81	0	81	0	5	81	4809	Iran Health Insurance Organization
	97	54	97	0	11	97	3670	mmunication Technology Organization of
	107	16	107	0	1	107	3258	The official newspaper of the country
	39	0	39	0	2	39	3163	Lake Urmia Rehabilitation Headquarters
	70	0	70	0	2	70	2815	stitute of Communication and Information
	52	0	52	0	6	52	2346	ousing and Urban Development Research
	46	0	46	0	10	46	2038	Isfahan Regional Water Company
	16	4	16	0	3	16	1757	Social Security Organization
	32	0	32	0	2	32	1729	Railway Research Center
	32	0	32	0	1	32	1585	/ice President for Science and Technolog
	20	0	20	0	1	20	1319	National mapping agency
	29	0	29	0	5	29	1307	Space Agency
	15	15	15	0	1	15	1264	zation of Municipalities and Villages of the
	10	0	10	0	1	10	1202	Literacy Movement Organization
	14	0	14	0	1	14	1114	Airports and Air Navigation Company
	13	0	13	0	1	13	1095	Corporate Audit
	24	1	24	0	3	24	1081	Nomadic Affairs Organization
	12	0	12	0	1	12	1066	Space Research Institute

Figure 5: View of the extracted data, the number of visits to each organization, and its data set.

records include the data extracted from each organization based on these criteria, which were extracted in the second phase. The "number of visits" attribute has been used as a label. Also, the names of government organizations and institutions are considered as ID. In this phase, the regression model is used to analyze the data and calculate the weight and importance of each criterion. The output of this model is the coefficients that determine the weight and importance of each criterion. Rapid Miner software has been used for this purpose. Figure 6 shows the operators used in this software.

The Split Data operator is used to divide the data into two sets of training and testing data. 70% of the data were considered as training data, and the remaining 30% as test data. Figure 7 shows the training data, including attributes, records, IDs, and labels. The number of training data was 78, which were randomly selected from a total of 112 available cases.

The linear regression operator is used to analyze the data and calculate the regression coefficients. Figure 8 shows the coefficients. The positive sign of the coefficient indicates the positive effect of the specified criterion on the label, which is the number of visits to the data set. The more positive and larger the coefficient of a criterion, the greater its impact on increasing the number of citizens visiting open government data. A negative sign indicates the negative impact of a specified criterion on the number of visits. The negative and larger the coefficient of a criterion, the greater its impact on reducing the number of visits to open government data.

As can be seen in Figure 8, the "society subject" criterion, with a coefficient of 72,564 and a positive sign, had the greatest impact on the label, which is an increase in the number of visits to open government data. This means that citizens were more interested in data sets that were relevant to the society. After that, the criterion of "format openness" with a coefficient of 52.682 and a positive sign has the second rank in increasing the number of visits. Therefore, citizens were more interested in data sets that can be read by a machine. Criteria for "metadata completeness," "number of data sets," "understandable," "data originality," "free," "lack of missing data," "nondiscriminatory," and "open license" being a positive sign, they gained the next ranks in increasing the number of visits to open government data. The important point is that the criterion of "number of categories of data sets" has a negative sign, which means that citizens are more inclined to data set that is more focused on a particular subject. Also, the "farming subject" with a negative sign and a coefficient of 160.413 had the most negative impact on the number of open government data views. This means that citizens were reluctant to visit the farm data set.

In order to evaluate the regression model and to evaluate the accuracy of the obtained coefficients, test data

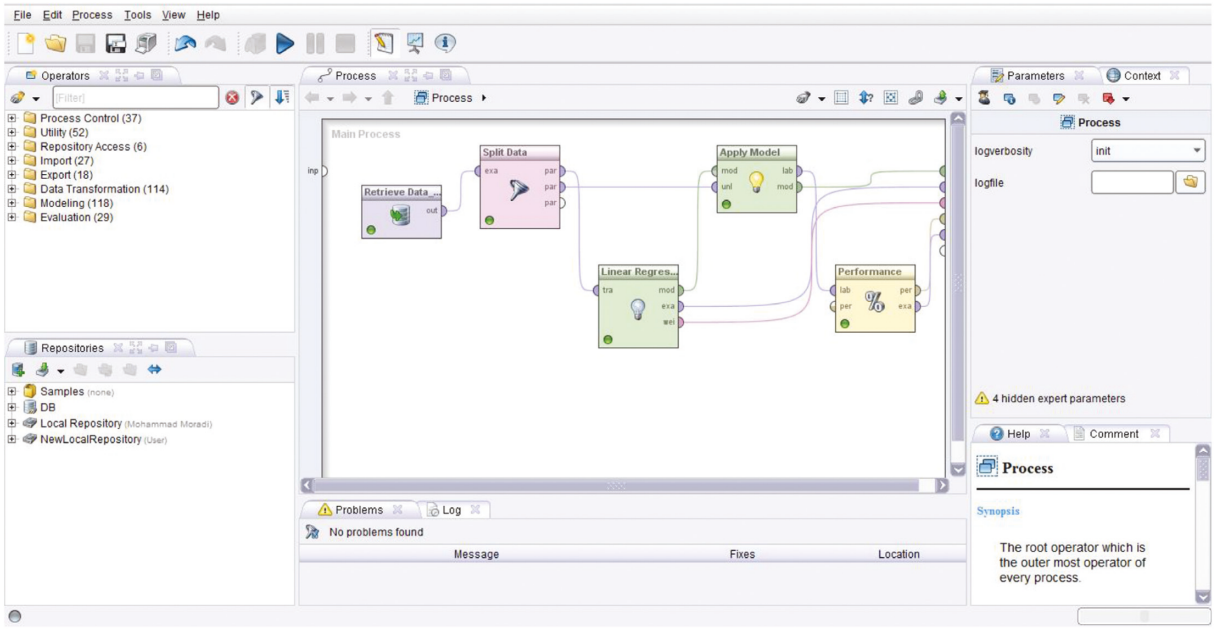


Figure 6: Operators used in Rapid Miner software to calculate regression coefficients.

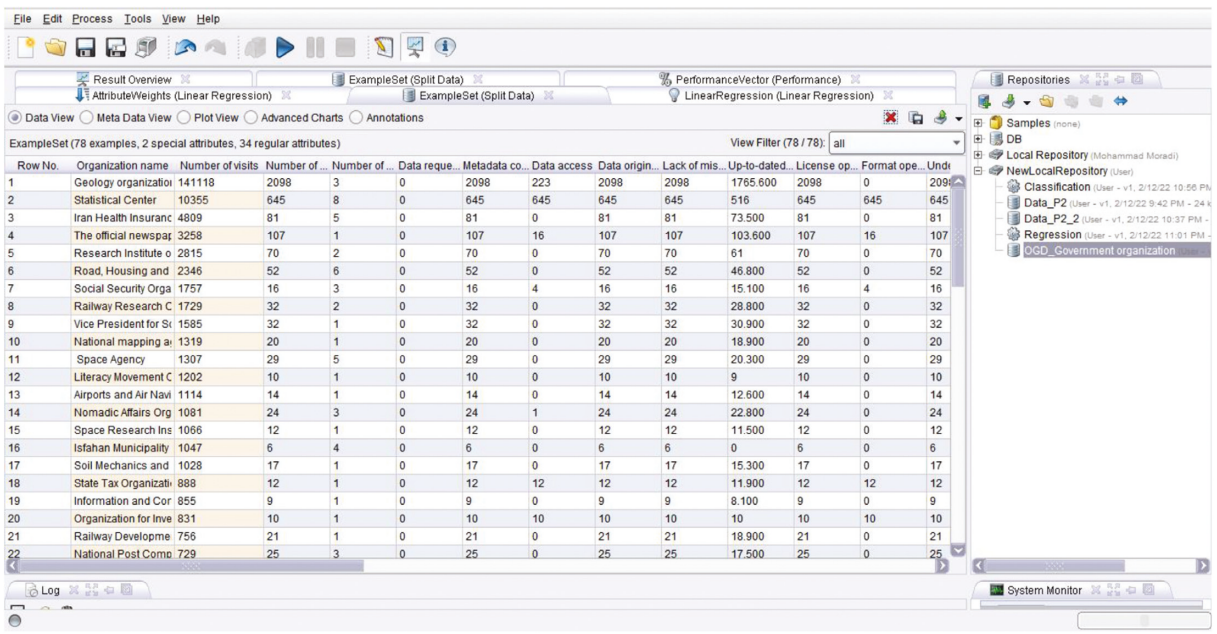


Figure 7: Training data.

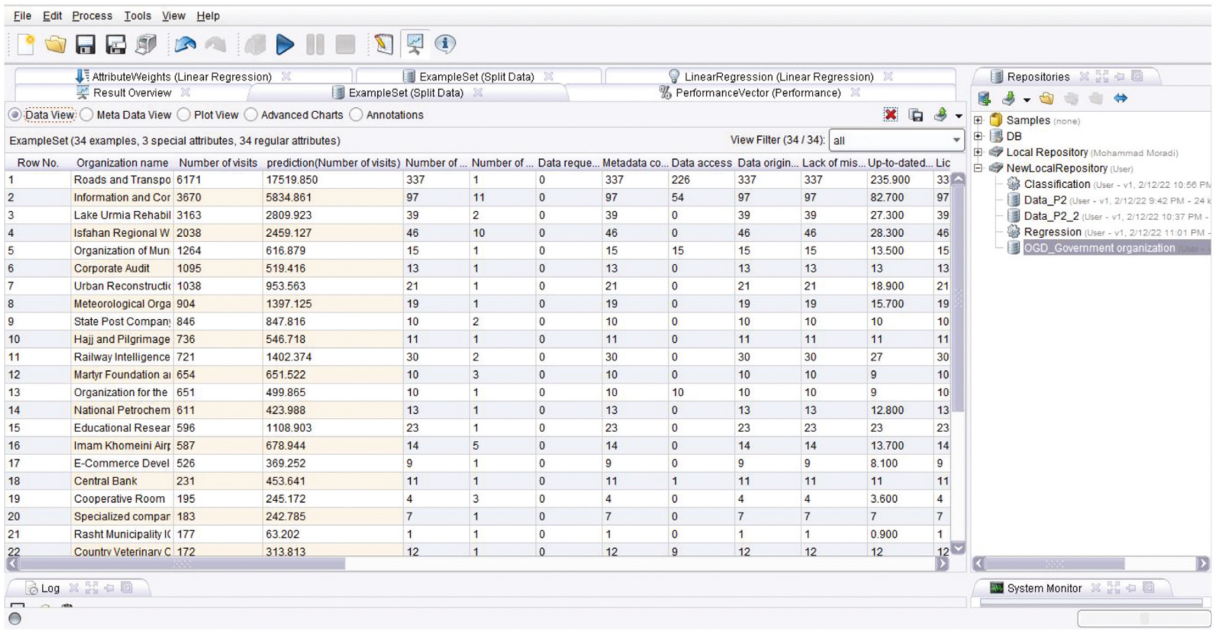
were used. 30% of the total data were used as test data. The number of test data was 78, which were randomly selected from a total of 112 items. Figure 9 shows the test data with the predicted rate for the label, which is the number of visits of open government data using the regression model generated. The actual number of visits is also shown.

Root-mean-square error (RMSE) was also used to evaluate the generated regression model, which was equal to 1998.435. According to the obtained RMSE, the value of normal root-mean-square error (NRMSE) was 0.014 (1.4%), which, according to [20] because it is less than 10%, indicates the desired state of the regression model.

4.5. Solutions. In this section, solutions to increase citizens' visits to organizations' data sets are stated. This section includes extracting solutions through library studies, as well as identifying the cause-and-effect relationships between the solutions based on the DEMATEL technique.

Attribute	Coefficient	Std. Error	Std. Coeffici...	Tolerance	t-Stat	p-Value	Code
Number of data sets	10.349	9.285	53.912	0.051	1.115	0.385	
Number of categories of data sets	-24.826	344.183	-21.768	0.979	-0.072	0.943	
Metadata completeness	11.196	9.285	58.323	0.051	1.206	0.322	
Data access	-43.375	7.264	-263.771	0.856	-5.971	0.000	****
Data originality	8.719	9.285	45.421	0.051	0.939	0.360	
Lack of missing data	8.306	9.285	43.268	0.051	0.895	0.383	
License openness	8.053	9.285	41.949	0.051	0.867	0.397	
Format openness	52.682	7.088	401.612	0.997	7.432	0	****
Understandable	8.773	9.285	45.699	0.051	0.945	0.357	
Free	8.402	9.285	43.767	0.051	0.905	0.377	
Non-discriminatory	8.085	9.285	42.117	0.051	0.871	0.395	
Borders subject	-77.242	245.458	-352.390	1.000	-0.315	0.757	
Images / Maps / Land cover	-12.094	167.598	-52.889	1.000	-0.072	0.943	
Location subject	-27.968	220.930	-145.792	1.000	-0.127	0.901	
Society subject	72.564	124.799	384.651	0.998	0.581	0.569	
Health subject	-9.888	71.144	-72.756	1.000	-0.139	0.891	
Management subject	-42.155	42.181	-270.373	1.000	-0.999	0.330	
Farming subject	-160.413	31.526	-1406.931	0.997	-5.088	0.000	****
Science and research education su	-25.032	80.365	-90.318	1.000	-0.311	0.760	
Energy subject	-41.681	83.672	-158.187	0.999	-0.498	0.625	
Structure subject	-59.139	22.011	-430.373	0.997	-2.687	0.011	**
Economy subject	-34.340	22.894	-203.815	0.997	-1.500	0.177	
Transportation subject	-26.228	93.489	-68.123	1.000	-0.281	0.783	
(Intercept)	56.208	∞	?	?	0	1	

Figure 8: Regression coefficients based on each criterion.



Row No.	Organization name	Number of visits	prediction(Number of visits)	Number of ...	Number of ...	Data reque...	Metadata co...	Data access	Data origin...	Lack of mis...	Up-to-dated...	Lic
1	Roads and Transpo	6171	17519.850	337	1	0	337	226	337	337	235.900	33
2	Information and Cor	3670	5834.861	97	11	0	97	54	97	97	82.700	97
3	Lake Urmia Rehabil	3163	2809.923	39	2	0	39	0	39	39	27.300	39
4	Isfahan Regional W	2038	2459.127	46	10	0	46	0	46	46	28.300	46
5	Organization of Mun	1264	616.879	15	1	0	15	15	15	15	13.500	15
6	Corporate Audit	1095	519.416	13	1	0	13	0	13	13	13	13
7	Urban Reconstructi	1038	953.563	21	1	0	21	0	21	21	18.900	21
8	Meteorological Orga	904	1397.125	19	1	0	19	0	19	19	15.700	19
9	State Post Compan	846	847.816	10	2	0	10	0	10	10	10	10
10	Hajj and Pilgrimage	736	546.718	11	1	0	11	0	11	11	11	11
11	Railway Intelligence	721	1402.374	30	2	0	30	0	30	30	27	30
12	Martyr Foundation a	654	651.522	10	3	0	10	0	10	10	9	10
13	Organization for the	651	499.865	10	1	0	10	10	10	10	9	10
14	National Petrochem	611	423.988	13	1	0	13	0	13	13	12.800	13
15	Educational Resear	596	1108.903	23	1	0	23	0	23	23	23	23
16	Imam Khomeini Airp	587	678.944	14	5	0	14	0	14	14	13.700	14
17	E-Commerce Devel	526	369.252	9	1	0	9	0	9	9	8.100	9
18	Central Bank	231	453.641	11	1	0	11	1	11	11	11	11
19	Cooperative Room	195	245.172	4	3	0	4	0	4	4	3.600	4
20	Specialized compar	183	242.785	7	1	0	7	0	7	7	7	7
21	Rasht Municipality I	177	63.202	1	1	0	1	0	1	1	0.900	1
22	Country Veterinary C	172	313.813	12	1	0	12	9	12	12	12	12

Figure 9: Test data with the predicted value of the label and the actual value of the label.

4.5.1. Extraction of Solutions. Most of the challenges, such as the lack of up-to-date data, the small number of data sets presented, the lack of items such as data visualization, etc., are that government organizations are reluctant to share their data. The following are solutions and incentives to encourage government organizations to share data.

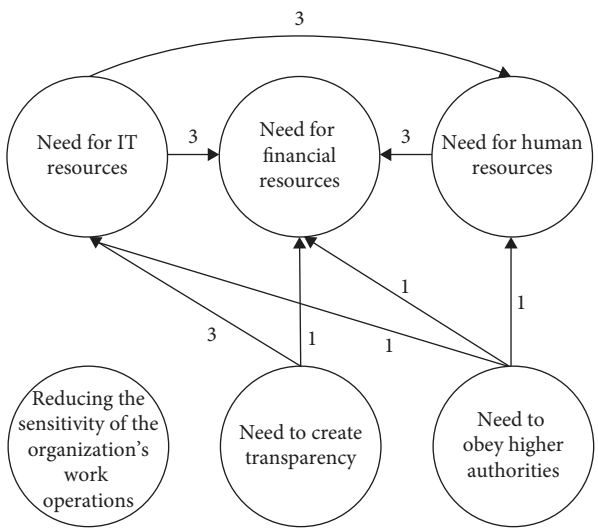


Figure 10: The direct-relation graph of solutions.

(1) Need for Human Resources [*21–23*]. External human resources, such as expertise and ideas, are vital to government organizations pursuing service innovation. Human resources include the manpower, ideas, knowledge, and skills needed to achieve the goals of the organization. Government organizations are looking for human resources and individuals to come up with ideas to help innovate public services. Therefore, it is necessary for government organizations to increasingly trust the expertise of foreign innovators to innovate services. Therefore,

(i) The relative need for human resources has a positive relationship with the dependence of the government organizations on foreign innovators. The greater this dependence, the more organizations will have to provide government data more openings to attract foreign innovators.

(2) Need for Financial Resources [*21*, *24*, *25*]. Financial resources here refer to the monetary capital needed by government organizations to achieve their goals of service innovation. Therefore, government organizations may rely on external sources for financial assistance. Therefore,

(i) The need for financial resources is positively related to the dependence of government organizations on external resources. The greater this dependence, the more organizations will have to provide more open government data to attract external financial resources.

(3) Need for IT Resources [*21*]. The use of IT resources for innovative activities in the services of government organizations is very important. For example, IT resources such as sensors and servers are essential for organizations seeking to innovate in smart city services. Government organizations may need external resources to provide IT resources for service innovation. Therefore,

(i) The relative need for IT resources has a positive relationship with the dependence of the government organizations on external resources. The dependence of government organizations on external sources is also positively related to the dissemination of open data.

(4) Need to Obey Higher Authorities [*21*, *26*]. One of the main factors in the involvement of government organizations in providing open data is higher authorities (for example, local and national governments) who implement formal and informal policies to influence organizations in data sharing. Such regulations or even the informal policies of higher authorities create institutional pressures on government organizations that shape their behavior. Therefore,

(i) The need for government organizations to follow higher authorities is positively related to its open data publishing behavior.

(5) Need to Create Transparency [*21*, *27*]. It is expected that government organizations' need for transparency will affect their open data behaviors. This is the freedom of information (FOI), which is recognized by the United Nations as a fundamental human right. When government organizations share their data with the public through open government data initiatives, citizens can control their activities and thus meet their need for transparency. Therefore,

(i) The need for transparency in government organizations is positively related to its open data publishing behavior.

(6) Reducing the Sensitivity of the Organization's Work Operations [*21*, *28–30*]. The specific work operations of a government organization affect its data-opening behavior. Some government organizations hold and process sensitive information because of their operations in government, which can restrict the provision of open government data. For example, government organizations related to health care need to process private information such as patients' medical records and background information. Similarly, government organizations working in national security, such as defense agencies, deal with limited information. Accordingly, the more sensitive the operations of organizations, the less data sharing will result. Therefore,

(i) The sensitivity of a government organization's operation is negatively related to its open data sharing behavior.

4.5.2. Analysis of Cause-and-Effect Relationships between Solutions Based on DEMATEL Technique. In this section, the solutions are examined based on the cause-and-effect relationships they have with each other. Identifying the cause-and-effect relationships between solutions will help managers invest more in solutions that have a greater impact on other solutions. The DEMATEL technique has been used for this purpose. Figure 10 shows the direct-relation graph of solutions based on the DEMATEL technique.

Table 2: The degree of causality and effectiveness between the solutions.

Solutions	R Degree of causality	J Degree of effectiveness	$(R+J)$ Sum of causality and effectiveness	$(R-J)$ $R-J>0 \longrightarrow$ definite cause $R-J<0 \longrightarrow$ definite effect
Need for human resources	0.5	1	1.5	−0.5
Need for financial resources	0	2.167	2.167	−2.167
Need for IT resources	1.25	0.667	1.917	0.583
Need to obey higher authorities	0.701667	0	0.701667	0.701667
Need to create transparency	1.291667	0	1.291667	1.291667
Reducing the sensitivity of the organization's work operations	0	0	0	0

Table 3: Ranking of solutions based on the degree of causality, the degree of effectiveness, and the degree of interaction with other factors.

Rank	Ranking based on the degree of causality	Ranking based on the degree of effectiveness	Ranking based on the degree of interaction with other factors
1	Need to create transparency	Need for financial resources	Need for financial resources
2	Need for IT resources	Need for human resources	Need for IT resources
3	Need to obey higher authorities	Need for IT resources	Need for human resources
4	Need for human resources	Need to obey higher authorities	Need to create transparency
5	Need for financial resources	Need to create transparency	Need to obey higher authorities
6	Reducing the sensitivity of the organization's work operations	Reducing the sensitivity of the organization's work operations	Reducing the sensitivity of the organization's work operations

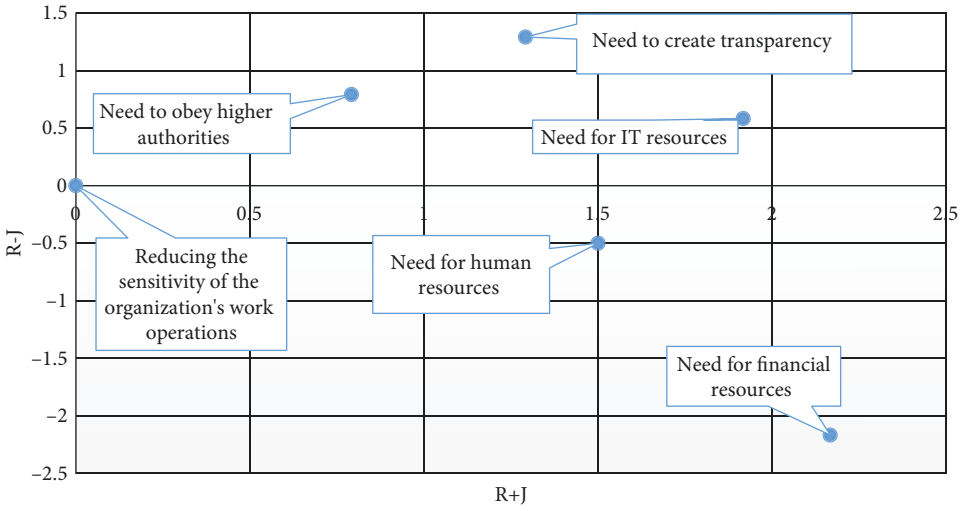


Figure 11: Impact chart between solutions.

The degree of causality and effectiveness between the solutions are shown in Table 2.

The ranking of solutions according to the degree of causality, the degree of effectiveness, and the degree of interaction with other factors is shown in Table 3.

According to Table 3, the "need to create transparency" solution has the greatest impact on other solutions. The "need for financial resources" solution is most affected by other factors. Also, the "need for financial resources" solution has the most interaction with other solutions. The $R-J$ is positive for "need for IT resources," "need to obey higher authorities," and "need to create transparency." Therefore, these solutions are causes in system. $R-J$ is also negative for "need for financial resources" and "need for human resources" solutions. Therefore, these solutions are the effects in the system. In Figure 11, by considering $(R+J)$ on the horizontal axis and $(R-J)$ on the vertical axis, the final position of the solutions in the system is specified. Factors above the axis $(R+J)$ are the causes and factors below the axis $(R+J)$ are the effects.

5. Conclusion

Despite the fact that some governmental organizations and institutions provided their data openly, the interaction of citizens with open government data was not favorable. This

issue can be caused by various factors such as the subject of the data set, the lack of format openness, the lack of up-to-datedness, and the lack of use of visual tools. Therefore, it is necessary to determine the effective criteria for the quality of open government data and increase the willingness of citizens to use the data. Unfortunately, each research has focused only on a specific dimension of open government data, and there is no comprehensive set of criteria. Also, the importance of each dimension and criterion is not considered. More importantly, the weight and importance of each criterion have not been calculated based on citizens' preferences to use different open government data. Also, the solutions and the importance of each of them have not been studied.

In this research, in the first phase, by studying and reviewing the articles, comprehensive and effective criteria were extracted in the quality of open government data and increasing the citizens' willingness from the data. The criteria extracted were included: "data originality," "license openness," "up-to-datedness," "data access," "metadata completeness," "number of data sets," "format openness," "nondiscriminatory," "understandable," "number of categories of data sets," "free," "lack of missing data," "data request ability," "visualization," "feedback," and "data subject matter." In the second phase, the data of 112 governmental organizations and institutions present in the open government data portal were extracted based on each of the stated criteria. In the third phase, in order to identify the organizations and their data sets that were most desired by citizens, the complex network of citizens and government organizations and institutions was analyzed. In order to identify organizations and government institutions with a high degree centrality, input links from citizens to organizations that represent the indegree of each organization were calculated. This information was available based on the number of visits to each organization's data set. In the fourth phase, data mining techniques including regression model were used to identify the data characteristics that were most desired by citizens. The output of the model was a coefficient that determined the positive or negative impact of each criterion as well as the weight and importance of that criterion. According to the results, the criterion of "society subject" with a coefficient of 72.564 and a positive sign had the greatest impact on increasing the number of citizens' visits to open government data. After that, the criterion of "format openness" with a coefficient of 52.682 and a positive sign has the second rank in increasing the number of visits. Criteria for "metadata completeness," "number of data sets," "understandable," "data originality," "free," "lack of missing data," "nondiscriminatory," and "license openness" being a positive sign gained the next ranks in increasing the number of citizens visiting open government data. The "number of categories of data sets" coefficient had a negative sign, meaning that citizens were more inclined to have a data set that focused more on a particular subject. Also, the "farming subject" with a negative sign and a coefficient of 160.413 had the most negative impact on the number of citizens' visits to open government data. Most of the challenges faced by higher ranking criteria such as format openness, small data sets, lack of data updates, etc., were that organizations were reluctant to present their data openly. Therefore, in the fifth phase, it was stated to provide solutions and incentives to increase the willingness of organizations to present their data openly. Based on the study and review of articles, six solutions include "need to create transparency," "need for IT resources," "need to obey higher authorities," "need for human resources," "need for financial resources," and "reducing the sensitivity of the organization's work operations" were extracted. Also, the cause-and-effect relationships between solutions were identified using the DEMATEL technique. Based on the results, the "need to create transparency" solution has the greatest impact on other solutions. The "need for financial resources" solution is most affected by other factors. Also, the "need for financial resources" solution has the most interaction with other solutions.

Extracting comprehensive and effective criteria in improving the quality of open government data and increasing citizens' willingness to use data, calculating the weight and importance of each criterion by analyzing the complex network of citizens and organizations, as well as providing solutions, can help managers make proper decisions and manage complex systems of citizens, government organizations, and institutions providing open government data in order to increase citizens' willingness to use the data and reap the benefits of open government data.

Data Availability

Data are available in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] K. Janssen, "Open government data and the right to information: opportunities and obstacles," *Journal of Community Informatics*, vol. 8, p. 2, 2012.

[2] M. Solar1, F. Daniels, and R. Lopez, "Automatic generation of roadmaps for open data," in *Electronic Government and Electronic Participation: Joint Proceedings of Ongoing Research, Posters, Workshop and Projects of IFIP EGOV 2014 and EPart*, vol. 21, p. 95, IOS Press, 2014.

[3] I. Susha, A. Grönlund, and M. Janssen, "Driving factors of service innovation using open government data: an exploratory study of entrepreneurs in two countries," *Information Polity*, vol. 20, no. 1, pp. 19–34, 2015.

[4] A. Halonen, "Being open about data," *Analysis of the UK Open Data Policies and Applicability of Open Data*, Finnish Institute in London, London, 2012.

[5] H. Yu and D. G. Robinson, "The new ambiguity of open government," *UCLA L. Rev. Discourse*, vol. 59, p. 178, 2011.

[6] A. Nikiforova and K. McBride, "Open government data portal usability: a user-centred usability analysis of 41 open government data portals," *Telematics and Informatics*, vol. 58, p. 101539, 2021.

[7] H. Zhang and J. Xiao, "Quality assessment framework for open government data: meta-synthesis of qualitative research,

2009-2019," *The Electronic Library*, vol. 38, no. 2, p. 1437, 2020.

[8] S. De Juana-Espinosa and S. ují-Mora, "Open government data portals in the European Union: a dataset from 2015 to 2017," *Data in Brief*, vol. 29, p. 105156, 2020.

[9] L. Zheng, W.-M. Kwok, V. Aquaro, X. Qi, and W. Lyu, "Evaluating global open government data: methods and status," *In Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*, pp. 381–391, 2020.

[10] K. Y. Dahbi, H. Lamharhar, and D. Chiadmi, "Toward an evaluation model for open government data portals," in *International Conference Europe Middle East & North Africa Information Systems and Technologies to Support Learning*, pp. 502–511, Springer, Cham, 2018.

[11] A. Vetrò, L. Canova, M. Torchiano, C. O. Minotas, R. Iemma, and F. Morando, "Open data quality measurement framework: definition and application to Open Government Data," *Government Information Quarterly*, vol. 33, no. 2, pp. 325–337, 2016.

[12] S. S. Dawes, L. Vidiasova, and O. Parkhimovich, "Planning and designing open government data programs: an ecosystem approach," *Government Information Quarterly*, vol. 33, no. 1, pp. 15–27, 2016.

[13] G. Misuraca and G. Viscusi, "Is open data enough? E-governance challenges for open government," *International Journal of Electronic Government Research*, vol. 10, no. 1, pp. 18–34, 2014.

[14] T. M. Harrison, T. A. Pardo, and M. Cook, "Creating open government ecosystems: a research and development agenda," *Future Internet*, vol. 4, no. 4, pp. 900–928, 2012.

[15] E. Estrada, *The Structure of Complex Networks: Theory and Applications*, Oxford University Press, United Kingdom, 2012.

[16] N. Veljković, S. Bogdanović-Dinić, and L. Stoimenov, "Benchmarking open government: an open data perspective," *Government Information Quarterly*, vol. 31, no. 2, pp. 278–290, 2014.

[17] R. Huang, C. Wang, X. Zhang, D. Wu, and Q. Xie, "Design, develop and evaluate an open government data platform: a user-centered approach," *The Electronic Library*, vol. 37, no. 3, p. 287, 2019.

[18] J. Attard, F. Orlandi, and S. Auer, "A systematic review of open government data initiatives," *Government Information Quarterly*, vol. 32, no. 4, pp. 399–418, 2015.

[19] S. Saxena, "Open government data (OGD) in six Middle East countries: an evaluation of the national open data portals," *Digital Policy, Regulation and Governance*, vol. 20, no. 4, p. 90014, 2018.

[20] G. Fu, *Modeling Water Availability and its Response to Climatic Change for the Spokane River Watershed*, Washington State University, United Status, 2005.

[21] Y. Zhenbin, A. Kankanhalli, S. Ha, and G. K. , "What drives public agencies to participate in open government data initiatives? An innovation resource perspective," *Information & Management*, vol. 57, no. 3, p. 103179, 2020.

[22] I. Mergel, "Opening government: designing open innovation processes to collaborate with external problem solvers," *Social Science Computer Review*, vol. 33, no. 5, pp. 599–612, 2015.

[23] D. Tapscott, A. D. Williams, and D. Herman, "Government 2.0: transforming government and governance for the twenty-first century," *New Paradigm*, vol. 1, p. l5, 2008.

[24] S. P. Taylor, "Innovation in the public sector: dimensions, processes, barriers and developing a fostering framework," *International Journal of Research Science & Management*, vol. 5, no. 1, pp. 28–37, 2018.

[25] S. P. Mohanty, U. Choppali, and E. Kougianos, "Everything you wanted to know about smart cities: the internet of things is the backbone," *IEEE Consumer Electronics Magazine*, vol. 5, no. 3, pp. 60–70, 2016.

[26] D. Zheng, J. Chen, L. Huang, and C. Zhang, "E-government adoption in public administration organizations: integrating institutional theory perspective and resource-based view," *European Journal of Information Systems*, vol. 22, no. 2, pp. 221–234, 2013.

[27] H. Yu and D. G. Robinson, "The new ambiguity of open government," *UCLA L. Rev. Discourse*, vol. 59, p. 178, 2011.

[28] A. Appari and M. E. Johnson, "Information security and privacy in healthcare: current state of research," *International Journal of Internet and Enterprise Management*, vol. 6, no. 4, pp. 279–314, 2010.

[29] S. E. Goodman and R. Ramer, "Global sourcing of IT services and information security: prudence before playing," *Communications of the Association for Information Systems*, vol. 20, p. 50, 2007.

[30] C. Coglianese, "The transparency president? The Obama administration and open government," *Governance*, vol. 22, pp. 529–544, 2009.