Proceedings of the 7th International Conference on Economics and Social Sciences (2024), ISSN 2704-6524, pp. 402-412

The 7th International Conference on Economics and Social Sciences
Exploring Global Perspectives:
The Future of Economics and Social Sciences
June 13-14, 2024
Bucharest University of Economic Studies, Romania

Towards a Robust Assessment Framework for the EU Open Data Maturity Index

Angelos FOUNTOULAKIS[1*], Anastasia PAPASTILIANOU[2]

DOI: 10.24818/ICESS/2024/040

Abstract

Strengthening Open Data policies is a priority for the Public Administrations of European countries because, firstly, they leverage their development momentum through economic, social, and environmental impacts and, secondly, they enhance transparency and accountability. In this context, the EU Commission publishes every year a report ranking a set of countries with respect to their maturity to Open Data. The methodology followed by the EU has given different results from year to year in terms of the ranking of a country, and the question that has been raised is the reliability of the method applied. The aim is to explore different decision methods to test their consistency and stability in relation to the existing method. In this study, a qualitative and quantitative research was conducted using a questionnaire and the AHP method was applied to modify the weights of the criteria that constitute the Open Data Maturity Index. Several countries show high volatility in their performance on sub-indices that are quite difficult to determine objectively. The AHP application showed that these sub-indices should have a fairly low weighting, having little impact on countries' performance. Based on the revised weights, the study arrives at a different ranking of the countries under evaluation and, combined with the use of the k-means method, a different clustering. A more structured and robust evaluation framework is proposed using ranking algorithms such as TOPSIS and PROMETHEE II.

Keywords: Open Data Maturity Index (ODMI), Multicriteria Decision Support Systems, Analytical Hierarchy Process, Data clustering, Public Administration.

JEL Classification: Z12.

[1] Hellenic Republic Ministry of Education, Religious Affairs and Sports, Athens, Greece; National School of Public Administration, Athens, Greece, aggelosfountoulakis@gmail.com.
* Corresponding author.
[2] National School of Public Administration, Athens, Greece, anastasiapapastilianou@gmail.com.

1. Introduction

According to the official open data portal (Open Data Portal, n.d.): "*The term Open (government) data refers to information collected, produced, or acquired for a fee by public bodies (another name is Public Sector Information) and made available free of charge for further use for any purpose*". As stated by Carsaniga et al. (2022), open data is a factor for social well-being because it enhances inclusiveness, financial transparency and accountability, as well as economic development. Characteristic cases are as follows: A) Austria, where an open data-based application has been developed that enables people with disabilities to know in real-time which lifts are operating in the Vienna metro. B) Bulgaria, which through the SEBRA system enables citizens to monitor in real-time public expenditure. C) Lithuania where an artificial intelligence application based on open data was developed, to address the problem of unemployment, by analysing the labour market, predicting local employment needs and proposing solutions.

Every year since 2015, the European Commission (2018-2022) has published a report, which initially aims to rank the countries of the European Union, some countries of the European Economic Area and some countries in pre-accession negotiations according to their maturity in Open Data. The Commission, from year to year, modifies the indicators to some extent. However, there are four general categories that, in principle, remain stable.

2. Problem Statement

The Commission's methodological framework relies on a simple averaging system that leaves countries exposed to fluctuations in their performance due to unweighted sub-indicators. Some sub-indicators are difficult to measure objectively and exhibit extremely high volatility both from year to year and between countries in the same-year reports. To address this issue, this paper proposes three main innovations in the way countries are scored and clustered. First, a multi-criteria scoring mode is adopted to assess not only overall performance, but also individual performance. Countries with high performance on highly weighted sub-indices compensate for any negative performance on low-weighted sub-indices, and bilateral comparisons of countries across all criteria shape the final ranking. Second, the weights of each sub-index are determined using the Analytic Hierarchy Process (AHP) method. The opinions of open data specialists and experts are incorporated into the weight of each index, reflecting a more realistic, mathematically based framework, like those found in the international literature, where indicators do not have equal weights. Third, the country clustering is performed using the k-means algorithm, which is widely used in machine learning. This method provides a more robust and reliable clustering framework that can accurately classify countries based on their open data maturity level.

AHP (Saaty, 1996) is based on the idea of hierarchy. Mishra et al. (2018) applied the AHP to evaluate factors that influence open data initiatives and actions in the Indian public administration. Schmid and Pape (2019) used the AHP to compare the security levels of information systems of different companies and to rank them in terms of their maturity. Arief et al. (2019) used the AHP to determine the importance of the business functions of the ICT department of the government in North Maluku. Kubler et al. (2018) used AHP to compare the quality of the metadata in 250 Open Data portals in 43 different countries.

The PROMETHEE algorithm (Brans & Mareschal, 1984) is used for finding the optimal solution to problems with many alternatives and many criteria. Panayiotou and Stavrou (2019) used the PROMETHEE II method to propose an evaluation framework for the maturity of e-services of the Greek Local Government. Balkan and Akyüz (2023) used the PROMETHEE method to assess the technological maturity of the countries belonging to the OECD. The advantage of PROMETHEE II is that it provides the possibility of a full comparison and ranking between alternatives, unlike PROMETHEE I.

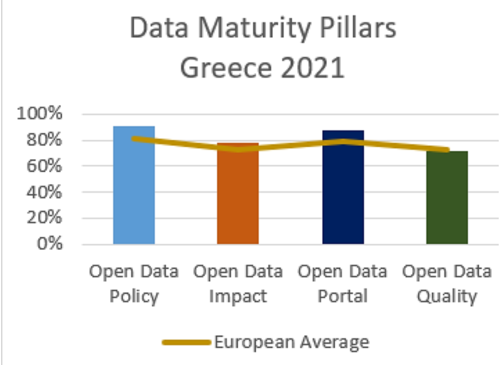
TOPSIS (Hwang & Yoon, 1981) is a multi-criteria decision analysis method (MCDM). TOPSIS presents a trade-off principle for multi-criteria decision-making processes. That is, there should be the smallest distance between the selected solution and the positive ideal solution and the largest distance between the selected solution and the negative ideal solution. Ardielli (2019) used TOPSIS to assess good governance in European Union countries. Sheoran et al. (2023) used the TOPSIS method to evaluate the usefulness and accessibility of open data portals, globally.

Pramanik et al. (2020) used the k-means algorithm, Haraty et al. (2015), to analyse data related to corruption in the public sector in Bangladesh. To the best of our knowledge, this is the first time that an integrated MCDM model combined with a clustering algorithm is used to modify the methodological framework of the EU Commission for the ODMI.

3. Research Questions / Aims of the Research

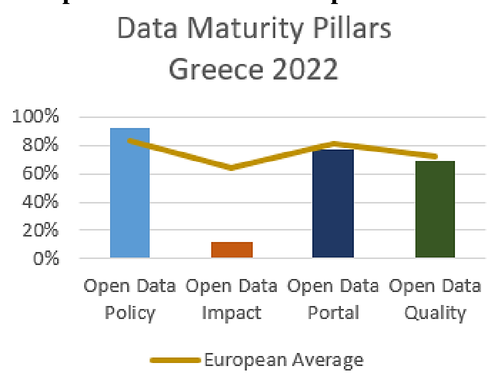
Qualitative research was conducted via Webex and in-person meetings with academics, experts in Decision Science, and government officials. The consultation and discussions lasted from 20th to 30th of June 2023 and took place in a total of four meetings. Based on the experts' suggestions, the Analytic Hierarchy Process (AHP) method was considered the most appropriate for determining the weights. For the quantitative research, a targeted sample of 60 people was selected based on eligibility criteria, including expertise and training on the subject of Open Data, previous experience, and the position held by the respondent. Although over 200 people were eligible to participate, the sample size was determined based on the eligibility criteria to ensure the quality of the responses. The questionnaire was compiled using Google Forms and sent to the selected respondents from July 5th to September 30th 2023 (responding rate 100%). The lion's share of the answers belongs to civil servants, because their involvement in consultations with the EU Commission and the PSI Group makes them perhaps the most appropriate to formulate the actual weight.

Figure 1. The performance of Greece per sub-indicator, 2021



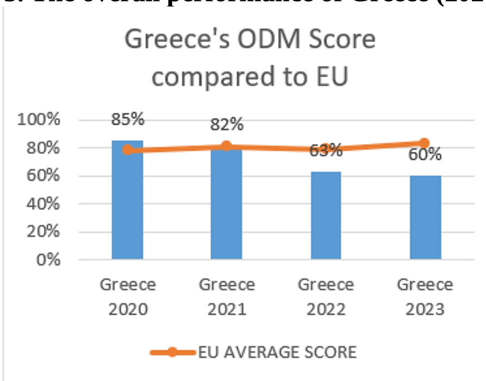
Source: processed by the authors, Excel.

Figure 2. The performance of Greece per sub-indicator, 2022



Source: processed by the authors, Excel.

Figure 3. The overall performance of Greece (2020-2023)



Source: processed by the authors, Excel.

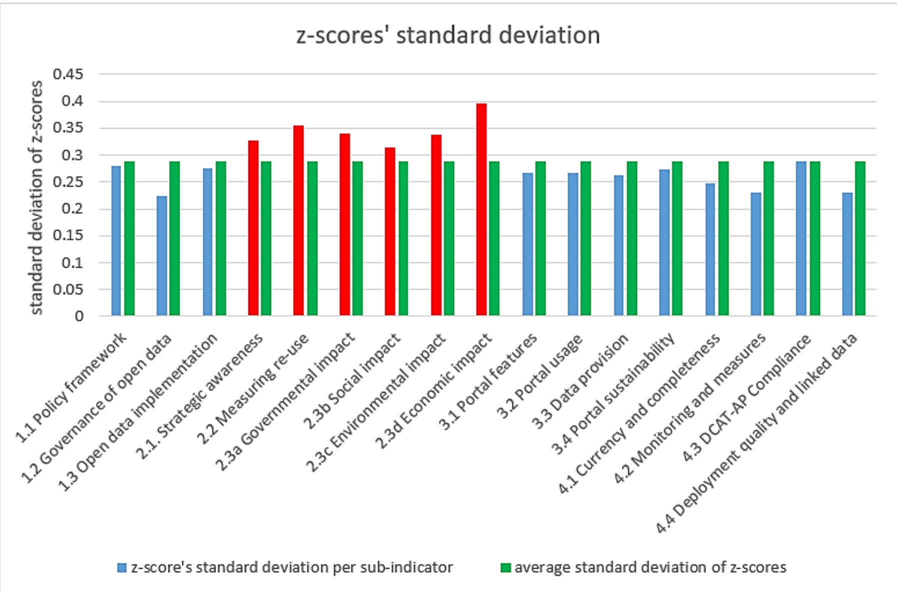
The questionnaire consists of a total of 114 sections. Greece's performance in three of the four indicators is stable (Figure 1 and Figure 2). However, from 2021 to 2022 there is a rapid deterioration in its performance in terms of “IMPACT”. Its inadequacy in this indicator is the cause of its overall score decline and its ranking in the lowest digital maturity tier. The overall score decreases from 82% in 2021 to 63% in 2022. A drop of 20% thanks to a single sub-indicator is unjustified in the context of an evaluation system that should be robust. The impact sub-index, as mentioned in the Open data Barometer 4th edition (World Wide Web Foundation, 2017), is difficult to measure and is prone to non-objective factors. The AHP method, gives high weight to the Policy sub-indicators and quite low weight to Created Impact. Greece's overall performance for 2019-2021 is equal to or above the EU average.

The Z-score, for normalisation Vafaei et al. (2016), in this study is defined as follows (in contrast to the mainstream z-score used in statistics):

$$z_{ij} = \frac{score_{ij} - \min_i\{score_{ij}\}}{\max_i\{score_{ij}\} - \min_i\{score_{ij}\}}\ ,\ i - country, j - criterion \quad (1)$$

For 2022 it is observed that in 6 of the 17 sub-indices the variability of the z-score is higher than the average variability. The problem is found in the sub-indicator (impact) with the Economic impact sub-dimension showing the highest volatility.

Figure 4. Z-scores volatility per indicator (2022)



Source: processed by the authors, Excel.

Table 1. The distribution of questionnaires

Academic Institutions	Public Sector	Private Sector
National Technical University of Athens, International University, National Kapodistrian University of Athens, The Higher School of, Pedagogical and Technological Education	Ministry of: a) Digital Governance, b) Development and Investments, c) The Interior, d)Tourism, e) Education, Religious Affairs and Sports, f) National Economy and Finance, Greek Parliament	Deloitte, Ernst & Young, MRC Energy Consultants, PFK Hellas

Source: authors' contribution.

4. Research Methods

Initially, the decision-makers express their view on the relative importance of the criteria, making bilateral comparisons. Their verbal response is converted into a whole number using the Likert scale, (1 = equally important, 2 = slightly more important, 3 = moderately more important, 4 = much more important, 5 = extremely more important). A matrix A is then formed, where each cell contains a number a_{ij} indicating the relative importance of the i criterion with respect to j. For example, if we use a 5-point Likert scale and the i criterion relative to j is considered much more important, then we would have $a_{ij} = 4$ and respectively $a_{ji} = \frac{1}{a_{ij}} = \frac{1}{4} = 0.25$. On the diagonal of the matrix the elements will be $a_{ii} = 1$, because the criterion i relative to itself is considered of equal importance. The weights of the criteria are obtained as the coordinates of the normalised eigenvector corresponding to the maximum eigenvalue:

$$Aw = \lambda_{max} w \ , \lambda_{max} \text{ (maximum eigenvalue)}, w \text{ (eigenvector)} \tag{2}$$

The Consistency Index is calculated as follows where n is the matrix A dimension:

$$CI = \frac{\lambda_{max} - n}{n-1} \tag{3}$$

The Consistency Ratio is defined as:

$$CR = \frac{CI}{RI} \tag{4}$$

where RI stands for Random Index and is a number given in official tables in the literature depending on the dimension of matrix A. If $CR < 10\%$ the results are accepted, otherwise the decision makers should repeat the procedure. For the TOPSIS method, the decision matrix is the table given by the EU Commission with the countries' scores in each sub-indicator, *Open Data Maturity Report (2022)* https://data.europa.eu/sites/default/files/country_scores_2022_0.xlsx, normalised with the Euclidean norm:

$$r_{ij} = \frac{y_{ij}}{\sqrt{\sum_{i=1}^{n} y_{ij}^2}}, \quad i = 1,2,\dots,n \ (countries) \ , \ j = 1,2,\dots m \ (criteria) \tag{5}$$

Calculation of the weighted normalised matrix:

$$w_{ij} = v_j r_{ij} \tag{6}$$

where v_j is the weight of criterion j. From the normalised weighted matrix W, the ideally optimal $H = (H_1,\ H_2, \dots, H_m)$ and the worst solution $D = (D_1,\ D_2, \dots, D_m)$ are calculated as follows:

$$H_j = \max_i(w_{ij})\ ,\ i = 1,2, \dots, n\ , j = 1,2, \dots, m \tag{7}$$

$$D_j = \min_i(w_{ij})\ ,\ i = 1,2, \dots, n\ , j = 1,2, \dots, m \tag{8}$$

The distance of each alternative from the ideally optimal and the worst-case solution is calculated:

$$d_i^+ = \sqrt{\sum_{j=1}^{k}(w_{ij} - H_j)^2} \quad d_i^- = \sqrt{\sum_{j=1}^{k}(w_{ij} - D_j)^2} \tag{9}$$

Calculate the relative distance of each alternative from the worst possible solution:

$$c_i = \frac{d_i^-}{d_i^- + d_i^+},\ \ i = 1,2, \dots, n \tag{10}$$

5. Findings

Table 2. The Consistency Ratio is below 10% for each of the six AHP implementations

AHP	Consistency Ratio (%)
Dimension Policy	5.2
Dimension Impact	<0.1
Dimension Portals	6.4
Sub-dimension Created Impact	4.3
Dimension Quality	7.1
Comparison of 4 main indicators	1.7

Source: data processed by the authors in Excel.

Table 3. The weights calculated by the AHP method for each indicator and sub-indices

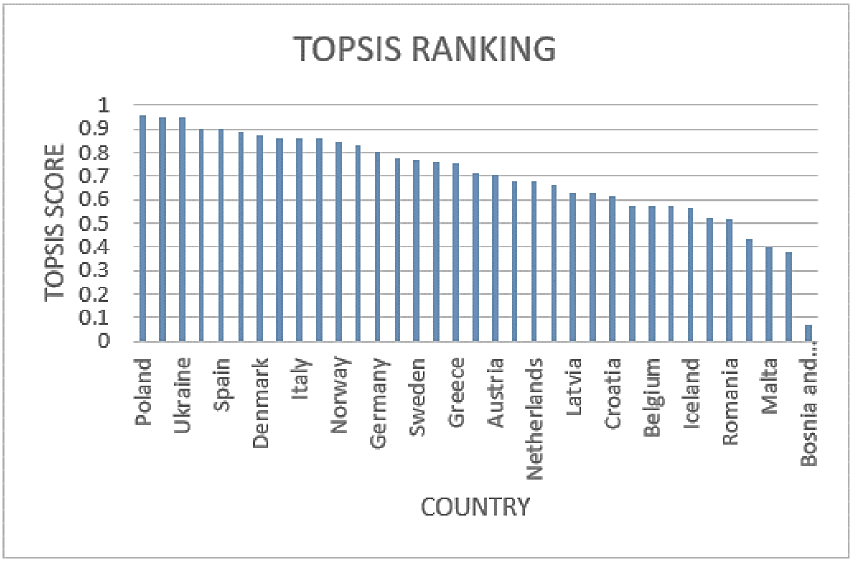
Policy 45.12%	Policy Framework		58.41%
	Governance of open data		27.65%
	Open data implementation		13.93%
Impact 10.15%	Strategic awareness		42.71%
	Measuring re-use		17.78%
	Created impact 39.51%	Governmental impact	14.89%
		Social impact	47.77%
		Environmental impact	10.32%
		Economic impact	27.02%
Portals 21.09%	Portal features		16.53%
	Portal usage		46.2%
	Data provision		26.86%
	Portal sustainability		10.41%
Quality 23.64%	Currency and Completeness		46.46%
	Monitoring and measures		27.89%
	DCAT-AP Compliance		15.85%
	Deployment quality and linked data		9.8%

Source: data processed by the authors in Excel.

For the implementation of PROMETHEE II we used the open-access software Visual PROMETHEE (Mareschal, 1988) https://bertrand.mareschal.web.ulb.be/promethee.html. Based on net flows ($\Phi(x_i) = \Phi^+(x_i) - \Phi^-(x_i)$, $\Phi^+(x_i)$: the sum of the advantages of one alternative over the others, $\Phi^-(x_i)$: the sum of the benefits of all other alternatives over the alternative under consideration), Greece is ranked 19th. It is worth noting that the first and last positions do not differ from the Commission's report for 2022. This is because the top countries have high scores on almost all indicators, so any differences in weighting leave them unaffected. The same is the case for the countries that are ranked last, such as Albania, Serbia, and Bosnia and Herzegovina, because they are still at an early pre-accession stage and, therefore, the necessary convergence has not yet taken place. The impact indicator, in which Greece lags behind, has a very low weighting. In contrast, the indicator Policy and Governance, in which Greece has scored above the European average, has an increased weighting.

The Commission's methodology with a ranking resulting from the weighted total score of each country, with AHP revised weights, ranks Greece 21st, contrary to the Commission's report of 2022 (28th). The one-dimensional version of the k-means method was applied, but the same results are obtained if we apply the multidimensional k-means where each vector will have 17 coordinates, i.e. the score of each country in each sub-index.

Figure 5. TOPSIS results (2022)



Source: data processed by the authors in Excel.

Table 4. The results of k-means clustering

Trend-Setters (AHP weights)	**Fast-trackers (AHP weights)**	**Followers (AHP weights)**	**Beginners (AHP weights)**
Ukraine, France, Poland, Ireland, Cyprus, Spain, Estonia, Italy, Slovenia, Denmark, Norway	Lithuania, Czech Republic, Germany, Finland, Sweden, Austria, Bulgaria, Hungary, Netherlands, Greece	Portugal, Croatia, Switzerland, Belgium, Latvia, Slovakia, Republic of Serbia, Romania, Luxembourg, Iceland	Montenegro, Malta, Albania, Bosnia and Herzegovina
Trend-Setters (equal weights)	Fast-trackers (equal weights)	Followers (equal weights)	Beginners (equal weights)
Ukraine, France, Poland, Ireland, Cyprus, Spain, Estonia, Italy, Slovenia, Denmark	Norway, Lithuania, Czech Republic, Germany	Finland, Sweden, Austria, Bulgaria, Hungary, Netherlands, Greece, Portugal, Croatia, Switzerland, Belgium	Latvia, Slovakia, Republic of Serbia, Romania, Luxembourg, Iceland, Montenegro, Malta, Albania, Bosnia and Herzegovina

Source: XLSTAT (software tool for k-means clustering) https://www.xlstat.com/en/.

6. Conclusions

The AHP results on the weight of the indicators are in perfect agreement with the variability of the z scores of each sub-indicator. Sub-indices with high variability receive less weight than sub-indices with lower variability of z score. Based on the new weights, the Policy indicator seems to be the most important, followed by the Portals and the Quality indicator, which are of approximately equal importance, and finally the Impact indicator. However, even with the equal-weights approach using the k-means algorithm, the ranking is different from that given by the EU Commission. The proposed methodological framework enhances long-term policy making. The three different approaches: PROMETHEE, TOPSIS, weighted average with adjusted weights indicate that Greece is ranked 19th-21st instead of 28th and place it among the Fast-trackers instead of the Beginners.

The study outlined above is subject to certain limitations that present opportunities for future research. The audience interview was limited to the Greek Public Administration. It is recommended that a pan-European survey be conducted with executives from all public administrations of the countries under evaluation. However, the results obtained seem to be in line with similar studies in the international literature, particularly when assessing the weights. Additionally, restructuring the clustering methodology using machine learning algorithms, setting up a special task force to assess the impact dimension, and implementing different multicriteria evaluation systems and sensitivity analysis will help achieve maximum robustness of the results.

Bibliography

[1] Arief, A., Natsir, D., Khairan, A., Sensuse, D.I. (2019). IT Governance Audit and Determination of Work Priorities using AHP: Case Study the Government of North Maluku, Indonesia. Journal of Physics: Conference Series, 1577, Article number 012046, DOI: 10.1088/1742-6596/1291/1/012066.

[2] Ardielli, E. (2019). Use of TOPSIS Method for Assessing of Good Governance in European Union Countries. Review of Economic Perspectives, 19(3), 211-231, DOI: 10.2478/revecp-2019-0027.

[3] Balkan, D., Akyüz, G.A. (2023). Technological maturity of the OECD countries: A multi-criteria decision-making approach using PROMETHEE. Cogent Engineering, 10(1), Article number 2219097, DOI: 10.1080/23311916.2022.2052485.

[4] Brans, J.P., Mareschal, B. (1984). PROMETHEE: A new family of outranking methods in multicriteria analysis. https://www.researchgate.net/publication/220700987_PROMETHEE_A_New_Family_of_Outranking_Methods_in_Multicriteria_Analysis.

[5] Carsaniga, G., Arriëns, E.N.L., Dogger, J., van Assen, M., Cecconi, G. (2022). Open Data Maturity Report 2022. Retrieved from https://data.europa.eu/sites/default/files/data.europa.eu_landscaping_insight_report_n8_2022_1_1.pdf; https://data.europa.eu/sites/default/files/country_scores_2022_0.xlsx.

[6] European Commission (2021). Open Data Maturity Report 2021. Retrieved from https://data.europa.eu/sites/default/files/landscaping_insight_report_n7_2021_0.pdf.

[7] Haraty, R.A., Dimishkieh, M., Masud, M. (2015). An Enhanced k-Means Clustering Algorithm for Pattern Discovery in Healthcare Data. Journal of Healthcare Engineering, Article number 615740. https://doi.org/10.1155/2015/615740.

[8] Hwang, C.L., Yoon, K. (1981). Multiple Attribute Decision Making: Methods and Applications. New York: Springer-Verlag, DOI: 10.1007/978-3-642-48318-9.

[9] Kubler, S., Robert, J., Neumaier, S., Umbrich, J., Le Traon, Y. (2018). Comparison of metadata quality in open data portals. Government Information Quarterly, 35 (1), 13-29, DOI: 10.1145/3132219.3132226.

[10] Mareschal, B. (1988). PROMETHEE. Retrieved from https://bertrand.mareschal.web.ulb.be/promethee.html.

[11] Mishra, A., Misra, D.P., Kar, A.K., Babbar, S., Biswas, S. (2018). Assessment of Open Government Data, Initiative-a Perception Driven Approach. 16th Conference on e-Business, eServices and e-Society (I3E), 159-171, DOI: 10.3389/fdata.2018.00002.

[12] Open Data Portal (n.d.). Open Data Portal. Retrieved from https://data.europa.eu/en.

[13] Panayiotou, N.A., Stavrou, V.P. (2019). A Proposed Maturity Assessment Framework of the Greek Local Government Web Electronic Services. Transforming Government People Process and Policy, 13(13/4), 237-256, DOI: 10.1108/tg-08-2018-0095.

[14] Pramanik, A. Sarker, A., Islam, Z., Hashem, M.M.A. (2020). Public Sector Corruption Analysis with Modified K-means Algorithm Using Perception Data. 11th International Conference on Electrical and Computer Engineering (ICECE), DOI: 10.1109/ICECE50432.2020.9368309.

[15] Saaty, T.L. (1996). Decision Making with Dependence and Feedback: The Analytic Network Process. Pittsburgh: RWS Publications.

[16] Schmid, M., Pape, S. (2019). A Structured Comparison of the Corporate Information Security Maturity Level. International Federation for Information Processing, DOI: 10.1007/978-3-030-26766-9_4.

[17] Sheoran, S., Mohanasundaram, S., Kasilingam, R., Vij, S. (2023). Usability and Accessibility of Open Government Data Portals of Countries Worldwide: An Application of TOPSIS and Entropy Weight Method. International Journal of Electronic Government Research, 19(1), 1-25, DOI: 10.4018/IJEGR.20230101.oa1.

[18] Vafaei, N., Ribeiro, R.A., Camarinha-Matos, L.M. (2016). Normalization Techniques for MultiCriteria Decision Making: Analytical Hierarchy Process Case Study. 7th Doctoral Conference on Computing, Electrical and Industrial Systems (DoCEIS), 261-269, ff10.1007/978-3-319-31165-4_26ff. ffhal-01438251f.

[19] World Wide Web Foundation (2017). Open data Barometer 4th edition. Retrieved from https://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf.