A multi-criteria decision making model for the selection of open data management systems   ISSN 1740-7494

Renáta Máchová

Faculty of Economics and Administration,
University of Pardubice,
Studentska 84, 53210 Pardubice, Czech Republic
E-mail: renata.machova@upce.cz

Martin Lněnička*

Department of Informatics
Business Academy and Secondary School of Tourism,
T. G. Masaryka 1000, 56501 Chocen
E-mail: martin.lnenicka@gmail.com
*Corresponding author

Abstract: A selection of advanced open and linked data technologies is very important in improving the impact of open government. This paper seeks to provide a multi-criteria decision model assessing open data management systems that are being deployed to enhance the e-government development. More precisely, it aims to provide a solution to the existing selection problem. The proposed evaluation framework consists of criteria identified through a systematic literature review focusing on selected open data management systems and their various quality dimensions. The fuzzy analytic hierarchy process was applied to validate the proposed model in an illustrative case study. For this purpose, an expert panel was established. Additionally, a sensitivity analysis was performed to understand the effects of the chosen priority weights. The results imply that open-source CKAN is the most suitable system (platform) that fulfils the requirements and characteristics of open data.

Keywords: open data management system; data quality; multi-criteria decision making; fuzzy analytic hierarchy process; case study.

Biographical notes: Renáta Máchová is with the Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Czech Republic as a Lecturer. Her work currently focuses mainly on issues relating to e-government development, open government and the quality of open data portals.

Martin Lněnička has received his Ph.D. in informatics in public administration from University of Pardubice, Czech Republic. His work currently focuses mainly on issues relating to e-government, big and open linked data, data lifecycle, and impacts of modern technologies in the public sector.

1 Introduction

Over the last few years, an increasing number of governments around the world have started to open up their data. This so-called open government movement has resulted in the launch of numerous open data portals and infrastructures aiming to provide a single point of access for government data and support increased public participation, collaboration, and cooperation (Drakopoulou, 2018; Hogan et al., 2017; Janssen, Charalabidis and Zuiderwijk, 2012; Máchová and Lněnička, 2017; Zuiderwijk, Janssen and Davis, 2014). Open government ecosystems then emphasize the multiple and varying interrelationships between data, open data providers, open data users, material infrastructures, and institutions (Janssen et al., 2012; Ndaguba and Ijeoma, 2018; Verma and Gupta, 2012; Zuiderwijk et al., 2014). Because of their scale, breadth, and status as the main source of information on a wide range of subjects, Open Government Data (OGD) are the most important driving force of open data progress in the society (Lněnička and Máchová, 2015; Van der Waal et al., 2014). Therefore, creating a portal of available datasets is a one way how to make OGD more accessible and easier

to find (Kučera, Chlapek and Nečaský, 2013). According to Hogan et al. (2017), these platforms have the potential to enhance transparency and trust in government.

Open data portals are an important part of the e-government infrastructure providing access to public services. They are powered by a data management system, which is used to publish, catalogue, search, and visualize open datasets (Braunschweig, Eberius, Thiele and Lehner, 2012; Kučera et al., 2013; Lnenicka, 2015). On the other hand, with the different data management systems powering open data portals, there is a great diversity in their functionalities, technologies, and implementations (Braunschweig et al., 2012; Máchová and Lněnička, 2017; Lourenço, 2015). According to Attard, Orlandi, Scerri and Auer (2015), usability is the most generic quality criterion, which means how easily can be the published data used. It is directly related to what degree open data are accessible, interoperable, complete, and discoverable. In this regard, the evaluation of these data management systems requires a careful review of related criteria and the selection of an appropriate decision making model focusing on their various quality dimensions. Several quality dimensions / perspectives have been introduced in the literature related to open data (Kubler, Robert, Le Traon, Umbrich and Neumaier, 2016; Umbrich, Neumaier and Polleres, 2015; Vetrò et al., 2016). This study considers some of those dimensions, and particularly the ones introduced in Lnenicka (2015), Máchová and Lněnička (2017), and Máchová, Hub and Lněnička (2018).

Furthermore, it is obvious that much real world knowledge is fuzzy rather than precise (Guitouni and Martel, 1998; Saaty and Vargas, 1987). Thus, there exist two kinds of concepts: clear concept and fuzzy concept. The clear concept refers to problems that are certain, definite, and specific. On the contrary, fuzzy concept refers to those problems that are uncertain, indefinite, and abstract (Hodgkin, Belton and Koulouri, 2005; Saaty and Vargas, 1987; Zadeh, 1975). A useful decision making model should be able to handle multiple fuzzy evaluations by aggregating the information of the related decision making problem (Chang, 1996; Cobo, Vanti and Rocha, 2014). A Fuzzy Analytic Hierarchy Process (FAHP) provides both the advantages of fuzziness, which is compatible with the uncertain, vague and subjective nature of decisions, and AHP as a systematic approach to make these decisions in complex Multi-Criteria Decision Making (MCDM) problems (Buckley, 1985; Chang, 1996; Saaty and Vargas, 1987).

Taking into account these problem areas, this paper aims to develop an evaluation framework and establish a decision making model that improves the open data management system evaluation process and enables stakeholders to choose the most suitable one. For this purpose, the framework consists of various quality dimensions that are integrated into a model using FAHP. From a practical standpoint, the proposed framework is applied to compare the most widely used open data management systems worldwide.

The structure of the paper is as follows: Problem formulation and research methodology are described in section 2. Literature review and theoretical background can be found in Section 3. Alternatives, criteria, and model description are provided in Section 4. In Section 5, a validation of the proposed model is conducted. The results are discussed in Section 6. Finally, conclusion remarks are given.

2   Problem formulation and research methodology

The FAHP method has been successfully applied to different fields in the e-government context. Markaki, Charilas and Askounis (2010) proposed an approach developing the evaluation framework that overcomes the problem of ambiguity by using triangular fuzzy numbers. They used a fuzzy MCDM approach to evaluate the quality features of e-government websites. The FAHP was implemented in order to offer a fuzzy weight for each features used in the selection procedure. A multi-criteria quality evaluation model for measuring the performance of e-government websites was also introduced by Kilinc (2010). In this hybrid model, the FAHP is used to evaluate the website quality analytically, while Fuzzy Inference System (FIS) is used to estimate priority order of projects according to the required effort and resource to improve the website quality. Syamsuddin and Hwang (2010)

developed a fuzzy MCDM framework to evaluate the security of e-government applications and infrastructures by modifying and improving the fuzzy hierarchical analysis by Buckley (1985). However, to the authors' best knowledge, there are no publications in the literature that evaluate open data management systems. Although Kubler et al. (2016) evaluated the quality of selected open data portals, they applied only the AHP method and compared data portals powered by the Comprehensive Knowledge Archive Network (CKAN) platform.

Since the environment where the new open data management system should be deployed can be unstable with lots of uncertainty, ambiguity, and uncertain factors involved, i.e., the progress of OGD portals and infrastructures implementation at the different administration levels, the FAHP is likely to be more suitable to address the selection problem. Therefore, the methodology of this paper follows the FAHP approach that is employed for the open data management system selection problem. After analysing the relevant literature and comparing various studies, an evaluation framework is proposed. It aims to present a systematic approach to the open data management systems evaluation and selection based on the linguistic evaluations with the fuzzy numbers. The framework is validated in an illustrative case study involving five domain experts. In contrast to the previous works applying the FAHP method in the e-government development context, this paper provides a solution to the existing selection problem regarding the specific quality requirements of open data management systems.

The criteria in this paper are assumed to be independent of all the others, because of the empirical evidence from a few similar studies (Kubler et al., 2016; Kučera et al., 2013; Vetrò et al., 2016). For the validation of the proposed model a combination of fuzzy extension of AHP by Chang (1996) and the geometric means method of Buckley (1985) and Buckley, Feuring and Hayashi (2001) is used. This approach is useful for obtaining a single assessment value based on different previously selected criteria and their sub-criteria. The relative importance of these evaluation criteria is modelled as a triangular fuzzy number. The FAHP is applied in order to provide a fuzzy weight for each criterion involved in the selection process. It is followed by the evaluation of identified alternatives. Crisp weights are obtained through the process of defuzzification. Finally, the CR is calculated and the consistency is checked. The main tools used are the MATLAB software and Microsoft Excel 2013.

2.1 Multi-criteria decision making and analytic hierarchy process

In general, the selection / evaluation of alternatives is a MCDM problem. In the past, many MCDM methods for this purpose have been developed (Guitouni and Martel, 1998; Hodgkin et al., 2005). The AHP is a MCDM technique founded by Saaty (1980) that integrates pairwise comparison ratios into a ratio scale. In classical AHP, the decision maker is asked to supply exactly pairwise comparison ratios for each criterion in each level of the hierarchy using a nine-point scale which converts the human preferences between available alternatives as equally, moderately, strongly, very strongly, or extremely preferred. Generally, it has three levels: goal, criteria, and alternative level, but additional levels can be created with sub-criteria, sub-sub-criteria, etc. The basic idea of typical AHP is based on the pairwise comparison matrices. One of the most common methods is the calculation of an eigenvector associated to the dominant eigenvalue of the comparison matrix. This value must be proximate to n and is also used to define the Consistency Index (CI) and the Consistency Ratio (CR), which imply that the estimate has logical consistency regarding the relative importance of the elements (Hodgkin et al., 2005; Saaty, 1980).

Though AHP is a convenient, flexible, and effective MCDM approach that combines qualitative analysis with quantitative analysis, it still has shortages in dealing with the transformation of qualitative information into quantitative information (Saaty and Vargas, 1987). More precisely, it has been generally criticized that it is not sufficient to take into account the uncertainty, ambiguity and uncertain factors involved in real-world decision problems associated with the mapping of human's preference judgments to a discrete scale (Kwong and Bai, 2003). As linguistic values are not mathematically operable (Zadeh, 1975), a logical way to overcome this limitation is to transform linguistic comparison ratios into fuzzy numbers. Finally, decision makers might be more reluctant to provide

crisp judgments than fuzzy ones (Kwong and Bai, 2003; Saaty, 1980). Hence, a number of methods have been developed to solve the situation where the comparison ratios are imprecise judgments and to generate weights such a fuzzy or interval comparison matrix (Buckley et al., 2001; Guitouni and Martel, 1998; Van Laarhoven and Pedrycz, 1983).

2.2 Fuzzy sets and fuzzy numbers in decision making

Zadeh (1965) firstly proposed the fuzzy set theory in 1965. In 1975, Zadeh (1975) further proposed and explored the linguistic variable whose values are words or sentences in a natural or artificial language. The earliest work in fuzzy AHP appeared in 1983 by Van Laarhoven and Pedrycz (1983) in which fuzzy ratios described by triangular membership functions were compared. Saaty and Vargas (1987) firstly presented interval judgments for the AHP method as a way to model subjective uncertainty. Fuzzy extension of the AHP method suggested by Chang (1996) then developed Saaty's AHP by integrating with fuzzy set theory to make decision making in uncertain environment. Later, Buckley et al. (2001) presented a new method of finding the fuzzy weights in fuzzy hierarchical analysis, which is the direct fuzzification of the original method used by Saaty in the AHP.

According to Dubois and Prade (1978) a fuzzy number is a fuzzy set (Zadeh, 1965) defined as $A = \{x, \mu_A(x)\}$ where x takes its number on the real line $\mathbb{R}$ and membership function $\mu_A: \mathbb{R} \rightarrow [0, 1]$, which have the following characteristics (Buckley, 1985; Dubois and Prade, 1978):

(1) A continuous mapping from $\mathbb{R}$ to the closed interval *[0, 1]*,

(2) Constant on *(-∞, a]*: $\mu_A(x) = 0 \; \forall x \in (-\infty, a]$,

(3) Strictly increasing on *[a, b]*,

(4) Constant on *[b, c]*: $\mu_A(x) = 1 \; \forall x \in [b, c]$,

(5) Strictly decreasing on *[c, d]*,

(6) Constant on *[d, ∞)*: $\mu_A(x) = 0 \; \forall x \in [d, \infty)$,

where *a, b, c, d* are real numbers and $a = -\infty$, or $b = c$, or $a = b$, or $c = d$ or $d = \infty$.

For convenience, μ_{LA} is named as left membership function of a fuzzy number A, defining $\mu_{LA}(x) = \mu_A(x)$, for all $x \in [a, b]$; μ_{RA} is named as right membership function of a fuzzy number A, defining $\mu_{RA}(x) = \mu_A(x)$, for all $x \in [c, d]$. For the sake of simplicity, trapezoidal or triangular fuzzy numbers are the most commonly used numbers (Buckley et al., 2001; Chang, 1996; Kwong and Bai, 2003). A triangular fuzzy number is a special type of fuzzy number whose membership is defined by three real numbers, expressed as (a_1, a_2, a_3), where a_1 is the lower limit, a_2 the most promising and a_3 the upper limit value (Cobo et al., 2014; Dubois and Prade, 1978; Zadeh, 1975). The membership function $\mu(x)$ of this triangular fuzzy number may therefore be described as (1) (Chang, 1996):

$$\mu(x) = \begin{cases} (x - a_1)/(a_2 - a_1), & x \in [a_1, a_2], \\ (a_3 - x)/(a_3 - a_2), & x \in [a_2, a_3], \\ 0, & otherwise \end{cases} \quad (1)$$

where $a_1 \leq a_2 \leq a_3$. If $a_1 = a_2 = a_3$ the fuzzy number gets a crisp number. It is possible to use the operation laws following Zadeh's extension principle via this simplification which makes calculations much easier. Given the triangular fuzzy numbers A = (a_1, a_2, a_3) and B = (b_1, b_2, b_3), the basic arithmetic operations are defined as follows (Dubois and Prade, 1978; Zadeh, 1975):

$$A+B = (a_1+b_1,\ a_2+b_2,\ a_3+b_3)$$

$$A*B = (a_1b_1,\ a_2b_2,\ a_3b_3)$$

$$nA = (na_1,\ na_2,\ na_3) \text{ for all } n>0$$

It is also possible to define the concepts of opposite $-A = (-a_3, -a_2, -a_1)$ and inverse $1/A = (1/a_3, 1/a_2, 1/a_1)$ fuzzy triangular numbers.

Using the concept of triangular fuzzy numbers, one can obtain a fuzzy or diffuse version of the classical AHP. When comparing two elements E_i and E_j, the exact value ratio a_{ij} can be approximated with a fuzzy ratio which is represented by a fuzzy triangular number. The construction of a hierarchical model in FAHP is exactly equal to the original AHP (Cobo et al., 2014). The fuzzy numbers required to form the decision matrix may be determined directly according to the decision maker or may derive from linguistic variables in a verbal scale, which can be then converted into fuzzy numbers using a suitable conversion as shown in Table 1 and the fuzzy pairwise comparison scale depicted in Figure 1. In order to construct a positive reciprocal matrix of pairwise comparisons, a full set of $n\,(n-1)\,/\,2$ comparison judgments are required. The pairwise comparison matrix is constructed as equation (2) (Cobo et al., 2014; Dubois and Prade, 1978):

$$\tilde{A} = \begin{pmatrix} \tilde{1} & \tilde{a}_{12} & ... & \tilde{a}_{1n} \\ \tilde{a}_{21} & \tilde{1} & ... & \tilde{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{a}_{n1} & \tilde{a}_{n2} & ... & \tilde{1} \end{pmatrix} \quad (2)$$

where $\tilde{a}_{ij} = \left(a_{1ij}, a_{2ij}, a_{3ij}\right)$, $\tilde{1} = (1, 1, 2)$ and $\tilde{a}_{ji} = 1/\tilde{a}_{ij}$. Finally, the concrete weights of the decision elements can be calculated using different methods that have been proposed in the literature, and can be found e.g. in Buckley et al. (2001), Guitouni and Martel (1998) or Hodgkin et al. (2005). One of the most popular methods is the Fuzzy Extent Analysis, proposed by Chang (1996). It consists of three steps and the final normalization weight vector *W* is a real non-fuzzy vector that can be used to perform an evaluation of the alternatives based on the weights of the decision elements in each hierarchy level. The local priorities represent the relative weights of criteria within a group with respect to their parent in the hierarchy. The global priorities are obtained by multiplying the local priorities of the siblings (sub-criteria) by their parent's local priorities (Chang, 1996; Cobo et al., 2014). In addition, the priorities may be also calculated via the geometric means method of Buckley (1985) and Buckley et al. (2001).

Table 1: Triangular fuzzy numbers to construct the pairwise comparison matrices (equivalence between the AHP and FAHP approaches)

Intensity of importance of one criterion over another	AHP Crisp number	FAHP	
		Triangular fuzzy scale	Fuzzy number
Equally important	1	(1, 1, 2)	A_1 or ($\tilde{1}$)
Weakly: moderately more important	3	(2, 3, 4)	A_3 or ($\tilde{3}$)
Essentially: significantly more important	5	(4, 5, 6)	A_5 or ($\tilde{5}$)
Strongly: strongly more important	7	(6, 7, 8)	A_7 or ($\tilde{7}$)
Absolutely: extremely more important	9	(8, 9, 9)	A_9 or ($\tilde{9}$)
Intermediate values	2; 4; 6; 8	(1, 2, 3); (3, 4, 5); (5, 6, 7); (7, 8, 9)	A_2; A_4; A_6; A_8

Source: Cobo et al. (2014), Dubois and Prade (1978)

Figure 1: The membership functions of the triangular fuzzy numbers for the FAHP
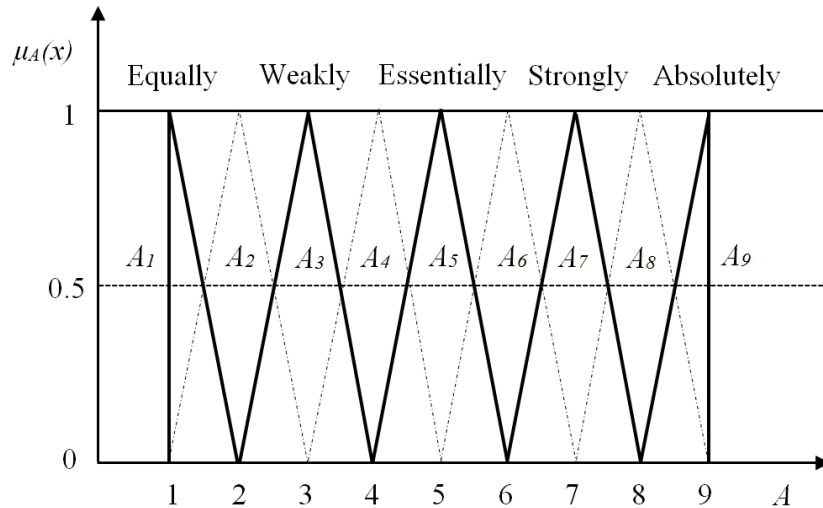
The triangular fuzzy numbers A_1, A_3, A_5, A_7 and A_9 are used to represent the pairwise comparison of decision linguistic variables from "Equally" to "Absolutely important than", and A_2, A_4, A_6 and A_8 represent the middle preference values between them.

The CI and CR of the pairwise comparison matrix must be considered whether bigger or smaller than 0.1. Let be a triangular fuzzy number A = (a_1, a_2, a_3). This fuzzy number can be defuzzied into a crisp value as equation (3) (Chang, 1996; Cobo et al., 2014):

$$P(A) = \frac{1}{6}(a_1 + 4 * a_2 + a_3) \tag{3}$$

The relative importance can be obtained by the right eigenvector w corresponding to the largest eigenvector (λ_{max}), in equation (4).

$$Aw = \lambda_{max}w \tag{4}$$

After defuzzification of each pairwise matrix, the CR for each matrix is calculated (Saaty and Vargas, 1987). This approach ensures that the FAHP is more efficient than the traditional AHP in a pair-wise comparison environment and provides better decisions due to use of linguistic values that can change from person to person. More precisely, since the traditional AHP cannot reflect the human thinking style, the FAHP was developed to solve the hierarchical fuzzy problems and provides less risky decisions.

3 Literature review and theoretical background

3.1 Moving from an open government towards open data portals

Since opening up government data is valuable not only for the public sector agencies and institutions, but also for citizens, businesses and other stakeholders, OGD should be a shared resource. In this regard, there should be collaboration and cooperation of various stakeholders in an open data ecosystem to facilitate the use of OGD (Attard et al., 2015; Drakopoulou, 2018; Ndaguba and Ijeoma, 2018; Ruijer et al., 2017; Zuiderwijk et al., 2014). Since data are central to this ecosystem, an ability to discover the relevant datasets is a prerequisite to unlocking the potential of open data (Attard et al., 2015; Máchová and Lněnička, 2017; Lourenço, 2015; Yang, Lin and Yu, 2015; Zuiderwijk et al., 2014). More precisely, Hogan et al. (2017) argued that the ability to search, filter, aggregate, visualise, modify, customise, and analyse data are central needs.

Creating a portal of available datasets is a one way how to make these data more accessible, intelligible, and useable (Kučera et al., 2013). In the last few years, governments have launched open data portals, specialized websites where datasets are published with high-quality metadata and organized

into a searchable catalogue (Hogan et al., 2017; Kubler et al., 2016; Lourenço, 2015; Ubaldi, 2013; Van der Waal et al., 2014; Vetrò et al., 2016). In these portals, open datasets are categorized according to their domains, providers, formats, and other properties for better accessibility (Yang et al., 2015). They also provide keyword search and various browsing interfaces to help users find relevant datasets and retrieve corresponding metadata about the publisher as well as content of datasets related to geography, jurisdiction, and time period of data (Lnenicka, 2015; Verma and Gupta, 2012). In addition, these portals need to fulfil certain legal, administrative as well as technical requirements (Braunschweig et al., 2012; Lourenço, 2015).

While the main implementations of open data initiatives are open data portals, there exist a number of different implementations with various characteristics (Attard et al., 2015). Kučera et al. (2013) divided them into groups based on the geographical coverage (administration level): local, regional, national, and international. Lněnička and Máchová (2015) then extended this classification by adding a new level of open data aggregators as a top-level category of data catalogues. Another categorization can be made based on the web paradigm, i.e., the traditional Web 1.0 paradigm or the more recent Web 2.0 paradigm (Alexopoulos, Zuiderwijk, Charapabidis, Loukis and Janssen, 2014). Based on the maturity of open data portals, Colpaert, Joye, Mechant, Mannens and Van de Walle (2013) proposed a five stages system representing the main function that the data portal is built or used for. The categorization starts with portals linking to various datasets and continues towards a metadata portal for both the datasets and the reuse of them. Finally, a data hub is set up where data become a common resource.

However, although the importance of open data portals in enhancing transparency and accountability in the public sector is recognized, there are still various barriers faced by practitioners and researchers. Hogan et al. (2017) summarized most of them and stated that these are related to limited organizational resources and budget, legislative challenges, poor information quality, lack of usability and technical issues. Janssen et al. (2012) emphasized the lack of explanation of the meaning of data, and the lack of knowledge to make sense of data. Martin, Foulonneau, Turki and Ihadjadene (2013) presented these categories of risks: governance, economic issues, licenses and legal frameworks, data characteristics, metadata, access, and skills. Ubaldi (2013) then reported that governmental interest appears to be on presenting data in a particular fashion, which distracts from the increasing provision to data users that they are really interested in.

Therefore, the efficient development of open data portals makes it necessary to evaluate their maturity systematic, in order to understand them better and provide the datasets and related features that are required by stakeholders. One solution is to define criteria meeting the quality aspects (Kubler et al., 2016; Kučera et al., 2013; Máchová and Lněnička, 2017; Umbrich et al., 2015; Vetrò et al., 2016).

3.2 Open data quality and related dimensions

A data quality is usually described in the literature by a series of quality dimensions that represent a set of consistency properties for a data artefact (Batini, Cappiello, Francalanci and Maurino, 2009; Kubler et al., 2016; Vetrò et al., 2016). Batini et al. (2009) published a detailed and systematic description of methodologies to assess and improve data quality. Methodologies are compared along several dimensions, including the methodological phases and steps, the strategies and techniques, the data quality dimensions, the types of data, and, finally, the types of information systems addressed by each methodology.

Measures on data quality can be applied in the open data domain (Vetrò et al., 2016; Yang et al., 2015). As reported by Kubler et al. (2016), there are existing issues with the quality of the metadata in data portals and the datasets themselves. This is a serious risk that could disrupt the open data project since the data quality needs to be managed to guarantee the reliability of public services. Moreover, as stated by Vetrò et al. (2016), it is necessary to consider that low-quality data provision increases the costs of accessing and interpreting data. Thus, open data domain-specific data quality criteria and measurements emerged recently to evaluate the quality of open data portals as well as

datasets and their metadata (Kubler et al., 2016; Lnenicka, 2015; Lněnička and Máchová, 2017; Maali, Cyganiak and Peristeras, 2010; Umbrich et al., 2015; Yang et al., 2015). In this regard, this issue has to be solved in the context of various quality dimensions that represent the most important characteristics of open data portals and related datasets (Kubler et al., 2016; Máchová and Lněnička, 2017; Umbrich et al., 2015; Vetrò et al., 2016; Zuiderwijk and Janssen, 2015).

According to the literature analysed, open data portals should address the whole range of entity types (thematic categories) and source-specific metadata content (Kubler et al., 2016; Lourenço, 2015). Further, the semantics and language that each open data portal is tied to, is one of the most common and inherent quality challenges (Petychakis, Vasileiou, Georgis, Mouzakitis and Psarras, 2014). Research on the quality of open data portals has confirmed that these portals should provide basic functionalities for uploading and downloading data (Alexopoulos et al., 2014; Charalabidis, Loukis and Alexopoulos, 2014). There should be also opportunities for data users to participate in improving published datasets (Alexopoulos et al., 2014; Zuiderwijk and Janssen, 2015). Specifically, the portals do not provide features for sharing and requesting information about datasets on social media platforms, thus limiting the potential reuse of these data (Ruijer et al., 2017). In addition, features gaining feedback are very limited, thus limiting data users from providing useful information to data providers in order to enhance the quality of data published online (Hogan et al., 2017). Specific quality dimensions were defined by Kubler et al. (2016), namely usage, completeness, openness, addressability, and retrievability. This paper extended their previous work in Umbrich et al. (2015). Kučera et al. (2013) defined these requirements of datasets quality: accuracy, completeness, consistency, and timeliness. Vetrò et al. (2016) proposed a framework of indicators to measure the quality of OGD on a series of data quality dimensions at most granular level of measurement.

4 Evaluation framework, alternatives, criteria and model description

Based on the literature review, a set of 3 dimensions (criteria) and 14 sub-criteria affecting open data management system selection is determined. It forms the evaluation framework in Table 2. The framework follows the perspective of quality dimensions and metrics defined by Batini et al. (2009) and furthermore used in Kubler et al. (2016), Ruijer et al. (2017); Umbrich et al. (2015), and Vetrò et al. (2016). In contrast to these studies, the following framework also covers emerging dimensions such as data sharing and requesting using social media or data visualization and analytics tools needed to explore datasets available. Six alternatives were identified based on the comparison of open data management systems (platforms) and their quality dimensions presented in Hogan et al. (2017), Kubler et al. (2016), Kučera et al. (2013), Lnenicka (2015), Maali et al. (2010), Máchová and Lněnička (2017), and Millette and Hosein (2016). The goal of this survey was ensuring that the framework reflects the reality of current platforms.

Table 2: Identification and description of the key quality requirements determining the selection process

Dimension	Sub-criteria	Description	Authors
Navigation and collaboration	Language versions	A portal offers more language versions to gain more users (attention).	Charalabidis et al. (2014), Petychakis et al. (2014)
	Search engine (filter)	A portal provides strong dataset search capabilities and selection tools using different criteria for browsing through categories and filters.	Charalabidis et al. (2014), Kučera et al. (2013), Lourenço (2015), Petychakis et al. (2014), Ubaldi (2013)
	User account	A portal supports user account creation in order to personalize views and information shown.	Charalabidis et al. (2014), Lnenicka (2015), Millette and Hosein (2016)
	Forum (feedback)	A portal provides an opportunity to submit feedback from users to providers and a forum to discuss and exchange ideas among users.	Charalabidis et al. (2014), Hogan et al. (2017), Petychakis et al. (2014), Ubaldi (2013), Zuiderwijk and Janssen (2015

|			
	Help (documentation)	A portal includes a high-quality of documentation and a help functionality to show how to use the portal.	Charalabidis et al. (2014), Hogan et al. (2017), Lnenicka (2015), Ubaldi (2013)
	Social media (share and request)	A portal is connected to a social media platform to create a social distribution channel for open datasets. It has a feature to request new datasets.	Hogan et al. (2017), Petychakis et al. (2014), Ruijer et al. (2017); Ubaldi (2013), Zuiderwijk and Janssen (2015)
Access and analysis	Full data API	A portal provides API for stakeholders to access data and develop applications using open datasets.	Kubler et al. (2016), Maali et al. (2010), Millette and Hosein (2016), Petychakis et al. (2014)
	Linked data access	A portal provides linked data access to connect open datasets using JSON or RDF format.	Maali et al. (2010), Ubaldi (2013), Van der Waal et al. (2014)
	Data visualization and analytics tools	A portal provides visualization and analytics capabilities to gain information about a dataset, e.g., in charts or visualizations in maps.	Alexopoulos et al. (2014), Hogan et al. (2017), Millette and Hosein (2016), Petychakis et al. (2014)
Datasets specifications	Thematic categories	A portal provides thematic categories of datasets to address the main topics covered by open datasets.	Kučera et al. (2013), Lnenicka (2015), Petychakis et al. (2014)
	Tags (keywords)	A portal distinguishes categories (themes) from tags (keywords). Same tags should be used to classify data of the same type and category.	Charalabidis et al. (2014), Maali et al. (2010), Kučera et al. (2013), Lourenço (2015)
	Dataset metadata (description)	A portal provides datasets together with their description and how and for what purpose they were collected.	Charalabidis et al. (2014), Kubler et al. (2016), Kučera et al. (2013), Maali et al. (2010), Ubaldi (2013)
	Open data license	A portal provides license information related to the use of the published datasets.	Kubler et al. (2016), Maali et al. (2010), Petychakis et al. (2014)
	User rating and comments	A portal provides capabilities allowing to collect user ratings and comments on a dataset.	Charalabidis et al. (2014), Ruijer et al. (2017); Zuiderwijk and Janssen (2015)

Source: Authors

The first alternative is CKAN, which is an open-source data management system that makes open data accessible by providing tools to streamline publishing, sharing, finding, and using data. The CKAN is aimed at data providers (national or local governments, businesses or other organizations) wanting to make their data open and available. It uses its internal model to store metadata about the different records and presents it on a web interface that allows users to browse and search these metadata. It also offers a powerful API that allows third-party applications and services to be built around it. The Drupal Knowledge Archive Network (DKAN) is an open-source open data management system (platform) with a full suite of cataloguing, publishing, searching, and visualization features that allows organizations to easily share data with the public. The DKAN takes a different approach by integrating open data catalogue features into an existing content management system. It also provides a user interface for many site management activities. The Open Government Platform (OGPL) is also an open-source platform to promote government transparency, accountability, and greater citizen engagement by making more government data, documents, tools, and processes publicly available.

Junar delivers the easy-to-use, cloud-based open data platform that enables businesses, governments, and academia to free their data to generate innovation and citizen participation and collaboration, maximize transparency and accountability, and enhance efficiency. Junar can either provide a

complete data catalogue or can provide data via the API to a separate user catalogue. The Prognoz data portal provides online access to monitor and analyse national and global datasets structured by industries and markets. Socrata's cloud-based solution allows government organizations to put their data online, make data-driven decisions, and share insights with citizens. It provides a service for data publishing, metadata management, data catalogue federation, and exposure of data (data can be published manually, or through dedicated APIs). One distinguishing feature of Socrata is that it allows users to create views and visualizations based on published data and save those for others to use.

Based on the identified components of the evaluation framework, a four-level hierarchy model is developed as illustrated in Figure 2. This model provides a clear hierarchy structure of the decomposed selection problem and it is prepared to be implemented with the MCDM methods.

Figure 2: A MCDM model representing the decomposed selection problem
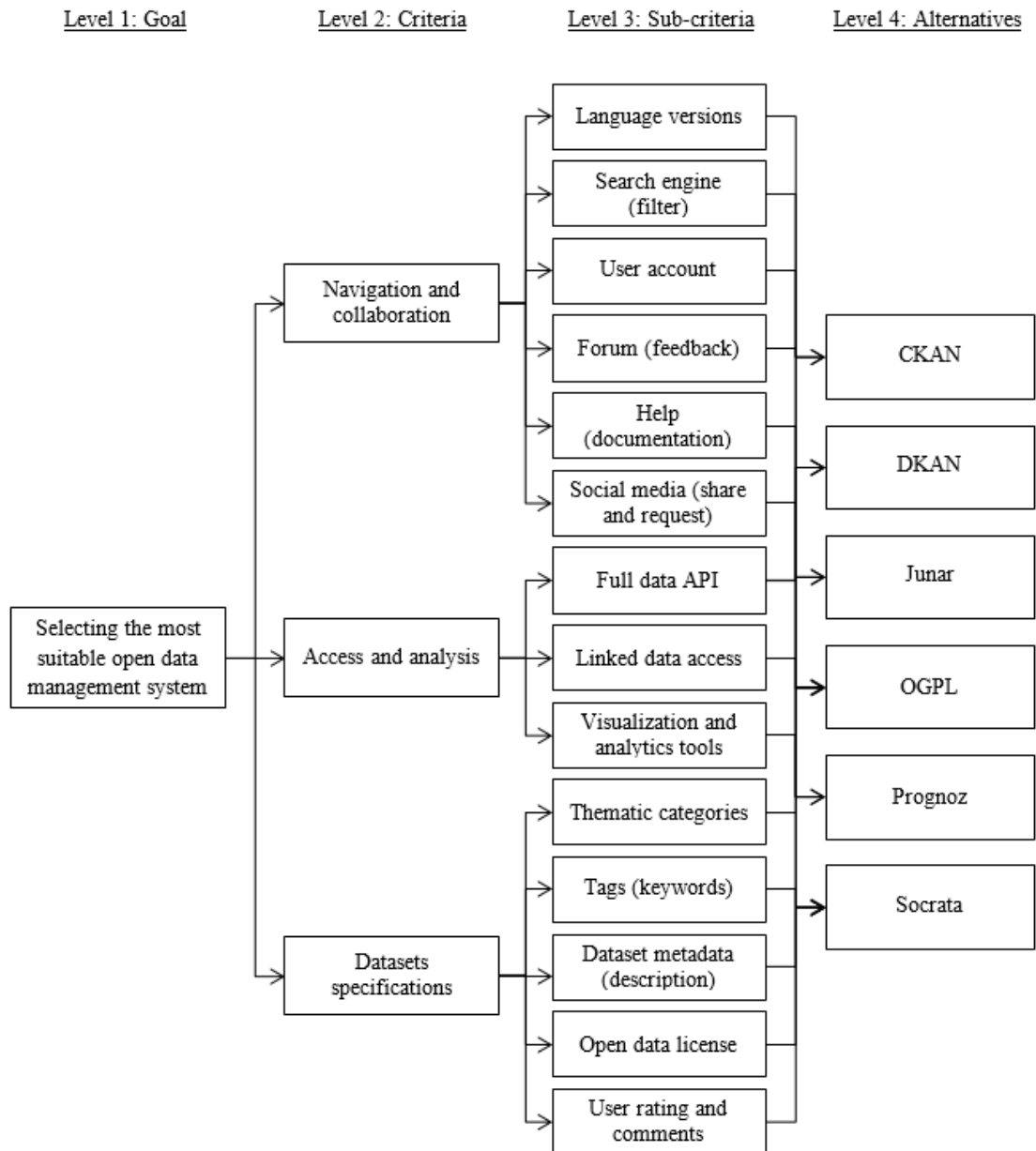
5   An illustrative case study

A real life situation is analysed in order to provide the better understanding, applicability, and validity of the proposed model. To validate the model, the FAHP approach is used. The following situation is focused on the selection of an open data management system suitable for the use on the national level as a central point for open data reuse. It should be cost-effective, easy to deploy and maintain, and support various open data licenses. For this purpose, five experts were asked to express their opinions (preferences) for each alternative with respect to each criterion. Since it is obvious that a difference in perception of this issue exists among various stakeholders, an expert opinion poll consisted of two researchers with several years of experience in e-government, OGD, and information systems; two senior public sector officials with more than 10 years of practice experience; and a person experienced in web design and development. The preferences of these experts had equal importance in decision making.

With respect to each alternative, the intensity of importance for each subjective criterion was rated on a linguistic term set (or linguistic scale) (Chang, 1996; Kwong and Bai, 2003). For this purpose, linguistics terms provided by fuzzy numbers in Table 1 are used, including intermediate values. Firstly, the pairwise comparisons for the set of criteria were performed and the matrices for each hierarchy level were obtained. These comparison matrices are constructed by using Saaty's scale but now with triangular numbers. The same was done for alternatives. Before all the calculations of vector of priorities, the comparison matrix has to be normalized into the range of [0, 1]. These steps were performed using the MATLAB software, including the geometric means method by Buckley (1985) and Buckley et al. (2001). Then, the fuzzy extension of AHP was used to compute the weights of criteria and sub-criteria. The result is called the final normalization weight vector *W* (Chang, 1996). However, this vector was used only for the criteria because it distinguishes between global and local priorities. The local priorities represent the relative weights of criteria within a group with respect to their parent in the hierarchy. The global priorities were obtained by multiplying the local priorities of the siblings by their parent's global priority. The global priorities for all the sub-criteria add up once again to 1 (Chang, 1996). Table 3 shows the triangular fuzzy numbers and summarizes both local and global weights for all the criteria. The crisp weights were obtained after defuzzification using equation (3).

Table 3: Local and global fuzzy and crisp weights for all the criteria

Hierarchy of criteria		Weights			
		Fuzzy		Crisp	
		Local	Global	Local	Global
1. criterion		**(0.217, 0.316, 0.545)**	-	**0.338**	-
	1.1	(0.064, 0.155, 0.217)	(0.008, 0.043, 0.112)	0.150	0.051
	1.2	(0.148, 0.287, 0.609)	(0.026, 0.085, 0.326)	0.318	0.107
	1.3	(0.041, 0.081, 0.186)	(0.003, 0.020, 0.095)	0.092	0.031
	1.4	(0.038, 0.079, 0.176)	(0.002, 0.019, 0.090)	0.088	0.030
	1.5	(0.091, 0.189, 0.398)	(0.014, 0.054, 0.211)	0.208	0.070
	1.6	(0.063, 0.151, 0.201)	(0.008, 0.042, 0.104)	0.145	0.049
2. criterion		**(0.145, 0.204, 0.378)**	-	**0.223**	-
	2.1	(0.105, 0.196, 0.272)	(0.009, 0.034, 0.097)	0.194	0.043
	2.2	(0.198, 0.277, 0.448)	(0.023, 0.051, 0.163)	0.292	0.065
	2.3	(0.251, 0.495, 0.854)	(0.030, 0.095, 0.317)	0.514	0.115
3. criterion		**(0.242, 0.425, 0.695)**	-	**0.439**	-
	3.1	(0.057, 0.136, 0.203)	(0.008, 0.052, 0.135)	0.134	0.059
	3.2	(0.053, 0.131, 0.194)	(0.007, 0.050, 0.129)	0.129	0.057
	3.3	(0.166, 0.317, 0.641)	(0.034, 0.129, 0.439)	0.346	0.152
	3.4	(0.151, 0.292, 0.625)	(0.031, 0.118, 0.428)	0.324	0.142
	3.5	(0.031, 0.058, 0.146)	(0.002, 0.019, 0.095)	0.068	0.030

Source: Authors

The geometric means method (Buckley, 1985; Buckley et al., 2001) was used for the evaluation of selected alternatives because it is much easier to use while working with more pairwise comparison matrices (Buckley et al., 2001). After defuzzification of each pairwise matrix, the CR for each matrix was calculated. At first, there was some inconsistency found, therefore, the judgments were redone until the CR was satisfactory. Table 4 shows the fuzzy and crisp weights of alternatives based on the geometric means calculations. There are also the five most important sub-criteria (with the highest priority) for each alternative. Fuzzy weights for alternatives are depicted as triangular fuzzy numbers in Figure 3.

Finally, the sensitivity analysis was performed to understand the impacts as a result of varying priority weights. For this purpose, preferences of each expert were separately calculated and compared to overall outcome. These results showed that datasets specifications was the most significant criterion for all the experts. Furthermore, the dataset metadata (description) was the most significant sub-criterion for three experts and the open data licence for two experts. Thus, the sensitivity analysis proved that the proposed model was almost reliable and stable, and only these two sub-criteria were replaced in priority levels for small changes in relative weights. No other significant changes in the rankings were observed, except for ranks of Junar, Prognoz, and Socrata, which may vary.

Table 4: Fuzzy and crisp weights for all the alternatives

Alternatives	Weights		Sub-criteria
	Fuzzy	Crisp	
CKAN	(0.168, 0.327, 0.715)	0.365	3.4, 3.3, 1.1, 2.3, 2.1
DKAN	(0.105, 0.189, 0.386)	0.208	3.4, 2.2, 3.3, 2.3, 1.2
Junar	(0.061, 0.081, 0.212)	0.100	3.3, 2.3, 1.2, 1.5, 3.1
OGPL	(0.079, 0.149, 0.221)	0.149	3.3, 3.4, 2.3, 1.2, 3.1
Prognoz	(0.038, 0.071, 0.201)	0.087	2.3, 1.2, 1.5, 3.1, 3.2
Socrata	(0.056, 0.075, 0.190)	0.091	3.3, 3.1, 2.3, 3.2, 1.2

Source: Authors

Figure 3: Fuzzy weights for all the alternatives
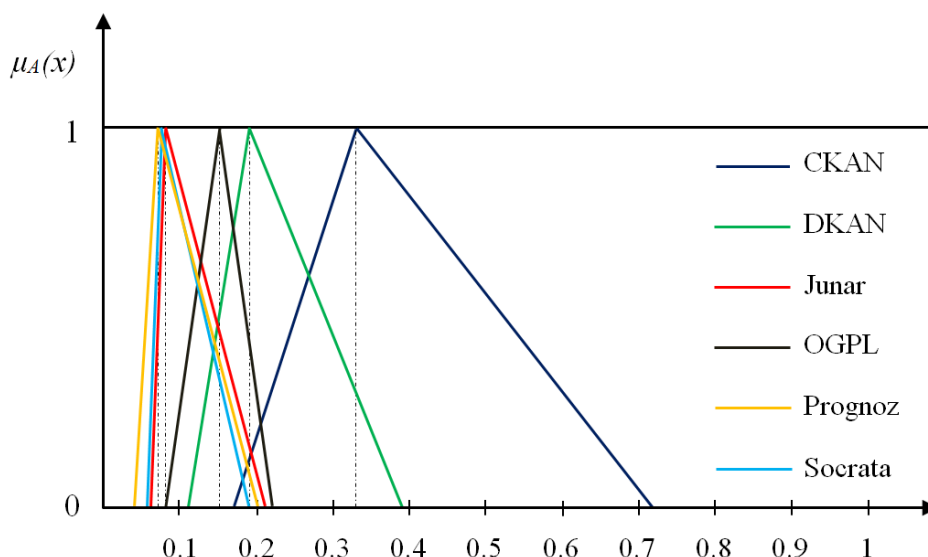
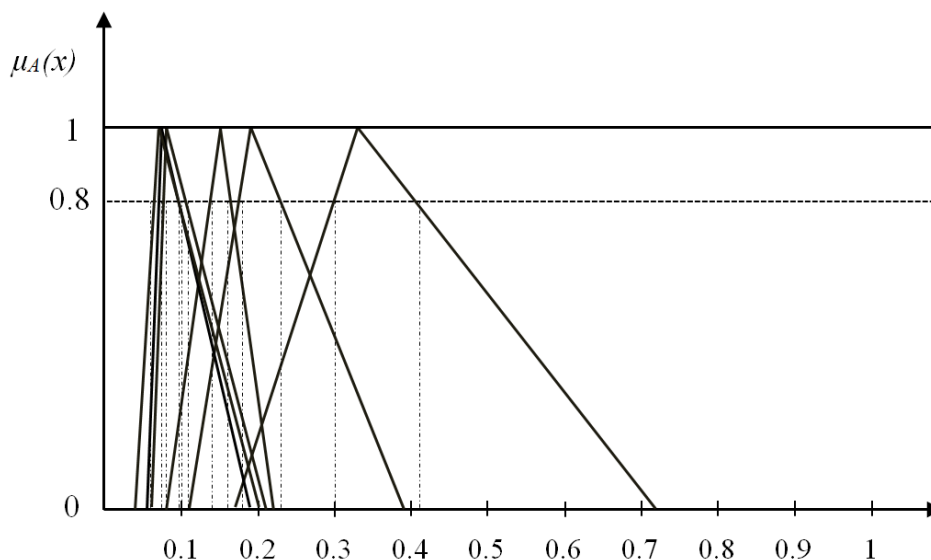*Source: Authors*

6 Results and discussion

As seen in the previous Table 3, the most important criterion affecting the selection of the open data management system is the description of a dataset and its metadata, i.e., title and short description, publisher, release date, and up to date of dataset, geographic coverage, and dataset URL. The quality of these metadata is crucial for open data providers to make their datasets available for reuse under the open data license, which is the second most important criterion. These are followed by visualization and analytics tools, which help stakeholders to gain a deeper understanding of the dataset before they use it.

As can be seen from the Table 4 and Figure 3, the most suitable open data management system is open-source CKAN. It is followed by the DKAN and OGPL. Finally, Junar, Prognoz and Socrata offer similar services, however, they focused mostly on data visualization and analytics tools. A closer linkage to open (government) data and licenses is missing. The sub-criteria with the highest priority

in each alternative's weight are usually the sub-criteria with the highest global weight (see Table 3), except of the DKAN that offers various features for linked data access, and Prognoz that allows to use advanced user account services.

The principal feasibility of fuzzy logic for the purpose of the MCDM is shown in Figure 4. The decision making process does not always imply a choice between alternatives, but could also refer to possibilities or considerations concerning opportunities (e.g. the rise of new technologies) or risks (e.g. restrictions and legal issues in the use of open data) in the e-government development. The fuzzy numbers could be taken to guarantee the minimum and maximum amount of costs for the promotion of open data reuse and investments in the open data management system. Therefore, in the following Figure 4, this is visualized for the membership function $\mu(x) = 0.8$, while a certain interval of related weights will be then calculated using equation (1): $0.298 \leq w_{CKAN} \leq 0.405$, $0.175 \leq w_{DKAN} \leq 0.229$, $0.076 \leq w_{Junar} \leq 0.107$, $0.137 \leq w_{OGPL} \leq 0.164$, $0.064 \leq w_{Prognoz} \leq 0.094$, $0.071 \leq w_{Socrata} \leq 0.096$.

Figure 4: Fuzzy evaluation of open data management systems with $\mu(x) = 0.8$

If the decision makers are certain with the information, the classical AHP method may be preferred and if information is not certain, the FAHP will tend to be the preferred option. It is worth noting that FAHP is suitable for the situations where the criteria are assumed to be independent of all the others, as was presented in this paper. Otherwise, the Analytic Network Process (ANP) method allows for complex interrelationships among decision levels and attributes. However, this method cannot be used with the proposed model, because of the quality dimensions hierarchy. Also, when a new alternative is added or an old one deleted in the decision making model, the sensitivity analysis should be done for finding out how it can affect the other alternatives and their ranks to each other.

There are also few challenges involved with our approach. First, while reflecting the limitations of the FAHP, there are external uncertainties referring to imprecise data, vague problem definition, etc., and internal uncertainties refer to subjectivity of the decision maker (Chang, 1996; Kwong and Bai, 2003). Further, considering the specific scenarios of usage and the specific needs of users, the requirements on open data portals may be different from the experts. In this regard, Hogan et al. (2017) argued that it is important to involve users in the development of the platform and be clear about what is what: when collected, by whom, how, etc. In this regard, a usability evaluation study using the AHP or the FAHP can be conducted to gain more insight into the current use of open data portals (Byun and Finnie, 2011; Máchová et al., 2018). Finally, various challenges of open data and existing datasets are limiting the potential possibilities of the whole ecosystem development and

should be explored more in future research (Drakopoulou, 2018; Ndaguba and Ijeoma, 2018). Particularly, the integration and implementation of these efforts in the broader context of e-government systems at the different administration levels is a challenging task (Al-Sebie, 2015).

7 Conclusions

The uncertainty, ambiguity, and uncertain factors involved in the environment generate another uncertainty for the decision makers. Therefore, it is more appropriate to develop a model with fuzzy numbers providing comprehensive decision making solution with ability to deal with inconsistent and vague judgments during the decision making process. This paper presents a new decision making model for the open data management system selection that is relatively new in the literature focusing on the e-government development. Five experts were involved in the study dealing with the validation of the proposed model. As resulted from the illustrative case study using the FAHP method and the sensitivity analysis, the evaluation framework as well as the four-level hierarchy model provide a base for the selection of the most suitable open data management system.

The selection of this system is a key factor in helping grow the availability and accessibility of open data across the Internet. Nowadays, these systems offer plenty of features and may be deployed as a multi-layered platform to fulfil various requirements of open and linked data ecosystems, such as the most suitable platform selected based on the new proposed model, the CKAN.

References

Al-Sebie, M. (2015) 'Technical challenges facing integration e-government systems: an empirical study', *Electronic Government, an International Journal*, Vol. 11, No. 3, pp.133–153.

Alexopoulos, C., Zuiderwijk, A., Charapabidis, Y., Loukis, E. and Janssen, M. (2014) 'Designing a Second Generation of Open Data Platforms: Integrating Open Data and Social Media', *Proceedings of the 13th IFIP WG 8.5 International Conference, EGOV 2014*, Springer, Berlin Heidelberg, pp.230–241.

Attard, J., Orlandi, F., Scerri, S. and Auer, S. (2015) 'A systematic review of open government data initiatives', *Government Information Quarterly*, Vol. 32, No. 4, pp.399–418.

Batini, C., Cappiello, C., Francalanci, C. and Maurino, A. (2009) 'Methodologies for data quality assessment and improvement', *ACM Computing Surveys (CSUR)*, Vol. 41, No. 3, pp.1–52.

Braunschweig, K., Eberius, J., Thiele, M. and Lehner, W. (2012) 'The State of Open Data – Limits of Current Open Data Platforms', *Proceedings of the 21st World Wide Web Conference 2012, Web Science Track at WWW'12*, ACM, New York, pp.1–6.

Buckley, J.J. (1985) 'Fuzzy Hierarchical Analysis', *Fuzzy Sets and Systems*, Vol. 17, No. 3, pp.233–247.

Buckley, J.J., Feuring, T. and Hayashi, Y. (2001) 'Fuzzy hierarchical analysis revisited', *European Journal of Operational Research*, Vol. 129, No. 1, pp.48–64.

Byun, D.H. and Finnie, G. (2011) 'An AHP method for evaluating usability of electronic government portals', *Electronic government, an international journal*, Vol. 8, No. 4, pp.343–362.

Chang, D.Y. (1996) 'Applications of the extent analysis method on fuzzy AHP', *European Journal of Operational Research*, Vol. 95, No. 3, pp.649–655.

Charalabidis, Y., Loukis, E. and Alexopoulos, C. (2014) 'Evaluating second generation open government data infrastructures using value models', *Proceedings of the 47th Hawaii International Conference on System Sciences*, IEEE, New York, pp.2114–2126.

Cobo, A., Vanti, A.A. and Rocha, R. (2014) 'A Fuzzy Multicriteria Approach for IT Governance Evaluation', *JISTEM – Journal of Information Systems and Technology Management*, Vol. 11, No. 2, pp.257–276.

Colpaert, P., Joye, S., Mechant, P., Mannens, E. and Van de Walle, R. (2013) 'The 5 stars of open data portals', *Proceedings of the 7th International Conference on Methodologies, Technologies and Tools Enabling e-Government, MeTTeG13*, University of Vigo, Vigo, pp.61–67.

Dubois, D. and Prade, H. (1978) 'Operations on fuzzy numbers', *International Journal of Systems Science*, Vol. 9, No. 6, pp.613–626.

Drakopoulou, S. (2018) 'Open data today and tomorrow: the present challenges and possibilities of open data', *International Journal of Electronic Governance*, Vol. 10, No. 2, pp.157–171

Guitouni, A. and Martel, J. M. (1998) 'Tentative guidelines to help choosing an appropriate MCDA method', *European Journal of Operational Research*, Vol. 109, No. 2, pp.501–521.

Hodgkin, J., Belton, V. and Koulouri, A. (2005) 'Supporting the intelligent MCDA user: A case study in multi-person multi-criteria decision support', *European Journal of Operational Research*, Vol. 160, No. 1, pp. 172–189.

Hogan, M., Ojo, A., Harney, O., Ruijer, E., Meijer, A., Andriessen, J., ... and Groff, J. (2017) 'Governance, Transparency and the Collaborative Design of Open Data Collaboration Platforms: understanding barriers, options, and needs', *Government 3.0–Next Generation Government Technology Infrastructure and Services*, Springer International Publishing, Cham, pp.299–332.

Janssen, M., Charalabidis, Y. and Zuiderwijk, A. (2012) 'Benefits, adoption barriers and myths of open data and open government', *Information Systems Management*, Vol. 29, No. 4, pp. 258–268.

Kilinc, M. S. (2010) 'Evaluation of e-Government Website Projects in Fuzzy Environment', *International Journal of Arts and Sciences*, Vol. 3, No. 9, pp.395–403.

Kubler, S., Robert, J., Le Traon, Y., Umbrich, J. and Neumaier, S. (2016) 'Open Data Portal Quality Comparison using AHP', *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*, ACM, New York, pp.397–407.

Kučera, J., Chlapek, D. and Nečaský, M. (2013) 'Open Government Data Catalogs: Current Approaches and Quality Perspective', *Technology-Enabled Innovation for Democracy, Government and Governance*, Springer, Berlin Heidelberg, pp.152–166.

Kwong, C.K. and Bai, H. (2003) 'Determining the importance weights for the customer requirements in QFD using a fuzzy AHP with an extent analysis approach', *IIE Transactions*, Vol. 35, No. 7, pp. 619–626.

Lnenicka, M. (2015) 'An In-Depth Analysis of Open Data Portals as an Emerging Public E-Service' *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, Vol. 9, No. 2, pp. 589–599.

Lněnička, M. and Máchová, R. (2015) 'Open (Big) Data and the Importance of Data Catalogs and Portals for the Public Sector', *Proceedings in Global Virtual Conference: The 3rd International Global Virtual Conference, GV-CONF 2015*, EDIS, Zilina, pp.143–148.

Lourenço, R.P. (2015) 'An analysis of open government portals: A perspective of transparency for accountability', *Government Information Quarterly*, Vol. 32, No. 3, pp.323–332.

Maali, F., Cyganiak, R. and Peristeras, V. (2010) 'Enabling interoperability of government data catalogues', *Proceedings of the 9th IFIP WG 8.5 International Conference, EGOV 2010*, Springer, Berlin Heidelberg, pp.339–350.

Máchová, R., Hub, M. and Lněnička, M. (2018) 'Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective', *Aslib Journal of Information Management*, Vol. 70, No. 3, pp.252–268.

Máchová, R. and Lněnička, M. (2017) 'Evaluating the Quality of Open Data Portals on the National Level', *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 12, No. 1, pp.21–41.

Markaki, O.I., Charilas, D.E. and Askounis, D. (2010) 'Application of Fuzzy Analytic Hierarchy Process to Evaluate the Quality of E-Government Web Sites', *2010 Developments in E-systems Engineering, DESE*, IEEE, New York, pp.219–224.

Martin, S., Foulonneau, M., Turki, S. and Ihadjadene, M. (2013) 'Open data: Barriers, risks and opportunities', *Proceedings of the 13th European Conference on eGovernment, ECEG 2013*, Academic Conferences and Publishing International Limited, Reading, pp.301–309.

Millette, C. and Hosein, P. (2016) 'A Consumer Focused Open Data Platform', *Proceedings of the 2016 3rd MEC International Conference on Big Data and Smart City, ICBDSC*, IEEE, New York, pp.1–6.

Ndaguba, E. and Ijeoma, E. (2018) 'Digital open government in Africa: a strategy for bridging citizens' gap', *Electronic Government, an International Journal*, Vol. 14, No. 2, pp.115–133.

Petychakis, M., Vasileiou, O., Georgis, C., Mouzakitis, S. and Psarras, J. (2014) 'A state-of-the-art analysis of the current public data landscape from a functional, semantic and technical perspective', *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 9, No. 2, pp.34–47.

Ruijer, E., Grimmelikhuijsen, S., Hogan, M., Enzerink, S., Ojo, A. and Meijer, A. (2017) 'Connecting societal issues, users and data. Scenario-based design of open data platforms', *Government Information Quarterly*, Vol. 34, No. 3, pp.470–480.

Saaty, T.L. (1980) *The Analytic Hierarchy Process*, McGraw-Hill, New York.

Saaty, T.L. and Vargas, L.G. (1987) 'Uncertainty and Rank Order in the Analytic Hierarchy Process', *European Journal of Operational Research*, Vol. 32, No. 1, pp.107–117.

Syamsuddin, I. and Hwang, J. (2010) 'A New Fuzzy MCDM Framework to Evaluate E-Government Security Strategy', *2010 4th International Conference on Application of Information and Communication Technologies, AICT*, IEEE, New York, pp.1–5.

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

Umbrich, J., Neumaier, S. and Polleres, A. (2015) 'Quality Assessment & Evolution of Open Data Portals', *Proceedings of the 3rd International Conference on Future Internet of Things and Cloud, FiCloud*, IEEE, New York, pp.404–411.

Van der Waal, S., Węcel, K., Ermilov, I., Janev, V., Milošević, U. and Wainwright, M. (2014) 'Lifting Open Data Portals to the Data Web', *Linked Open Data – Creating Knowledge Out of Interlinked Data*, Springer International Publishing, Cham, pp.175–195.

Van Laarhoven, P.J.M. and Pedrycz, W. (1983) 'A Fuzzy Extension of Saaty's Priority Theory', *Fuzzy Sets and Systems*, Vol. 11. Nos. 1-3, pp.229–241.

Verma, N. and Gupta, M. P. (2012) 'Open Government Data: More than Eighty Formats', *Proceedings of the 9th International Conference on E-Governance, ICEG 2012*, CSI, Cochin, India, pp.207–216.

Vetrò, A., Canova, L., Torchiano, M., Minotas, C.O., Iemma, R. and Morando, F. (2016) 'Open data quality measurement framework: Definition and application to Open Government Data', *Government Information Quarterly*, Vol. 33, No. 2, pp.325–337.

Yang, H.C., Lin, C.S. and Yu, P.H. (2015) 'Toward Automatic Assessment of the Categorization Structure of Open Data Portals', *Multidisciplinary Social Networks Research,* Springer, Berlin Heidelberg, pp.372–380.

Zadeh, L.A. (1965) 'Fuzzy sets', *Information and Control*, Vol. 8, No. 3, pp. 338–353.

Zadeh, L.A. (1975) 'The concept of a linguistic variable and its application to approximate reasoning – I', *Information Sciences*, Vol. 8, No. 3, pp. 199–249.

Zuiderwijk, A. and Janssen, M. (2015) 'Participation and data quality in open data use: Open data infrastructures evaluated', *Proceedings of the 15th European Conference on E-Government 2015, ECEG 2015*, Academic Conferences and Publishing International Limited, Reading, pp.351–359.

Zuiderwijk, A., Janssen, M. and Davis, C. (2014) 'Innovation with open data: Essential elements of open data ecosystems', *Information Polity*, Vol. 19, Nos. 1, 2, pp.17–33.