Computer Science and Information Systems 13(1):237–258 DOI: 10.2298/CSIS151210001A

A Methodology for Determining the Value Generation Mechanism and the Improvement Priorities of Open Government Data Systems

Charalampos Alexopoulos, Euripides Loukis, Yannis Charalabidis

University of the Aegean, GR-83200, Karlovassi, Samos
{alexop, eloukis, yannisx}@aegean.gr

Abstract. Government agencies all over the world are making big investments for developing information systems that open important data they possess to the society, in order to be used for scientific, commercial and political purposes. It is important to understand what value they create and how, and at the same time – since this is a relatively new type of information systems (IS) – to identify the main improvements they require. This paper contributes in this direction by presenting a methodology for determining the value generation mechanism of open government data (OGD) systems and also priorities for their improvement. It is based on the estimation of a 'value model' of the OGD system under evaluation from users' ratings, which consists of several value dimensions and their corresponding value measures, organized in three 'value layers', and also the relations among them. The proposed methodology has been successfully applied to an OGD system developed as part of the European project ENGAGE ('An Infrastructure for Open, Linked Governmental Data Provision towards Research Communities and Citizens'), and provided interesting insights and improvement priorities. This first application provides evidence that our methodology can be a useful decision support tool for important ODG systems investment, management and improvement decisions.

Keywords: open government data, public sector information, evaluation, decision support system, value model.

1. Introduction

In many countries all over the world there is an increasing interest in moving towards the 'Open Government' paradigm, which according to the 'Open Government Directive' of USA [12] has three main components: transparency (promoting accountability by providing the public with information about what the government is doing), participation (allowing members of the public to contribute ideas and expertise, so that their government can benefit from information and knowledge that is widely dispersed in society, in order to design better policies), and collaboration (improving the effectiveness of government by encouraging partnerships and cooperation within the federal government, across levels of government, and between the government and private institutions). Opening various types of data possessed by government to the citizens is a central element of the first of the above components, and also of critical importance for the achievement of the other two. As open data can be defined data that are freely accessible online, available without technical restrictions to re-use, and provided under open access license that allows the data to be re-used without limitation (for commercial and non-commercial purposes) [26]. One important category of open data are the Open Government Data (OGD), defined as "data and information produced or commissioned by government or government controlled entities" [26], which are opened up for use and re-use by public and private agents alike. In this direction government agencies worldwide are making big investments for developing information systems that open through the Internet important data they possess (e.g. concerning government activity and expenses, economic activity, businesses, health, pollution, traffic, unemployment, crime, poverty data, etc.) to the society, in order to be used for scientific, commercial and political purposes [2][4] [5] [6] [9] [18] [25] [35]. This constitutes an important shift of the public sector: from an information gatekeeper it gradually becomes an information provider as well, making its own contribution to the development of the information economy and society [8]. It is increasingly recognized that proactively opening public data has a good potential to create considerable benefits for several stakeholders, such as scientists, journalists and active citizens who want to understand better various public problems and policies through advanced processing of such data, or even firms and individuals interested in the development of value added e-services or mobile applications combining various types of government data (OGD), and possibly other private data. According to Jetzek [22] there are two 'ideologies' concerning the value potential of OGD: the first of them focuses of the economic value that can be generated from OGD (e.g. contribution to the development of new e-services and mobile applications which are traded in markets, having worth determined by these markets), while the second one focuses on the social value from OGD (e.g. contribution to improvements in the lives of individuals or society as a whole through better government policies, and in general to the generation of 'public value' not traded in markets).

Taking into account the abovementioned big investments made by numerous government agencies all over the world for the development of OGD infrastructures, and also the above expectations for important benefits and value generation from OGD, it is important to assess the value that these first OGD infrastructures really generate (i.e. to what extent the above expectations are realised), to understand better the various types of value generated and also their generation mechanisms, and at the same time – since this is a relatively new type of information systems – to identify the main improvements they require. However, there has been quite limited activity in this direction. A recent study of the OECD on OGD initiatives by Ubaldi [35] concludes that *'So far, little has been done to analyse and prove the impact and accrued value of these initiatives'*, and calls for action in this direction. It also notes that an important barrier for this is the lack of a structured and comprehensive evaluation methodology. Our review of previous relevant literature (see section 2) has concluded that though there are some methodologies for evaluating OGD initiatives at the level of country or individual government agency, there is a lack of methodologies for evaluating OGD systems, which is the most critical level for value generation from OGD.

This paper contributes to filling this research gap: it describes and validates a methodology for evaluating OGD infrastructures, which adopts the 'value model' estimation approach to IS evaluation proposed in [27] and [23]. According to this approach the evaluation of IS should include not only the assessment of various measures of generated value (as in the 'conventional' IS evaluation approaches), but also the relations among them as well, leading to the formation of a value model of the IS; this provides highly important advantages: it enables a deeper understanding of the whole IS value generation mechanism and also a rational definition of IS improvement priorities (see section 2 for more details on this approach). In particular, the proposed methodology is based on the estimation of a three layers' value model of an OGD infrastructure from users' ratings, following the value model approach to evaluation proposed in [27] and [23]. Its first layer includes measures of the value associated with the quality of the open data provision related capabilities it offers. Its second layer includes measures of the value associated with the support it provides to users for achieving their open data use related objectives. Finally, its third layer includes measures of the value associated with users' future behaviour with respect to the OGD infrastructure. For each of the above layers particular value measures are defined taking into account previous relevant IS research (see section 2) and also the particular capabilities that these OGD infrastructures offer. Furthermore, our methodology includes the estimation and exploitation of not only the magnitudes of all the above types of value generated by the OGD infrastructure, but also (going beyond the 'conventional' IS evaluation approaches) of the relations among them as well (which are neglected by the conventional approaches). This leads to the formation of a value model of the OGD infrastructure, which shows how capabilities related value lead to the creation of objectives' achievement support related value, and finally to the creation of users' future behaviour related value. Also, this value models allows identifying the capabilities of the OGD infrastructure (at the first layer of the value model) that should take the highest improvement priority. The proposed methodology has been used for the evaluation of an OGD infrastructure developed in the European project ENGAGE (for more details see http://www.engagedata.eu/about/).

In the following section 2 previous related research work is reviewed, and then in section 3 the theoretical background of the proposed methodology is outlined. In section 4 the proposed methodology is described, while in section 5 the abovementioned application of it is presented. Finally in section 6 the conclusions are summarized and future research directions are proposed.

2. Related Research Work

Some methodologies have been developed for evaluating OGD initiatives at the level mainly of country and of individual government agency, however there is a lack of methodologies for evaluating OGD systems, though this is the most critical level for value generation from OGD.

Several international organizations have developed methodologies/frameworks for evaluating and comparing the progress and maturity of countries with respect to government data opening. The Open Knowledge Foundation has developed the 'Open Data Index' (ODI) (see https://index.okfn.org/about), which aims to measure to what extent countries all over the world have released government data in a way that they are easily accessible to citizens, media and civil society. It proposes a number of assessment dimensions, which concern publicly available data, freely available data, data available online, data in machine readable formats, data available in bulk, up-to-date data, open licenses, available terms of use, metadata and data quality. The World

Bank has developed the 'Open Data Readiness Assessment' (ODRA) [40], which aims to support the assessment from various perspectives of the readiness of a national, or even a regional or municipal government, with respect to opening of government data. Its main evaluation dimensions are: leadership, policy/legal framework, institutional structure and capabilities, data management and availability, demand for open data, citizen engagement, open data ecosystem, funding, national technology infrastructure, and national skills infrastructure. The abovementioned study of the OECD on the OGD initiatives of its member countries [35] proposes a framework for the collection of data about and analysis of country-level OGD initiatives, which includes three main assessment dimensions: strategy and legal-institutional framework, implementation framework (with respect to technology, finance, organization, communication and ecosystem development), and impact and value creation (social, political and economic). The World Wide Web Foundation developed the 'Open Data Barometer' (ODB) [41], which has a wider scope, aiming to access countries' OGD readiness of all stakeholders, progress in implementation, and impact. In particular, the main evaluation dimensions it proposes are: government readiness, civil society and citizens' readiness, business and entrepreneurs' readiness; availability of data sets; data sets for innovation, data sets for social policy, data sets for accountability; political impacts, economic impacts and social impacts.

Furthermore, there has been some research towards the development of methodologies frameworks for the evaluation of OGD initiatives of individual government agencies. the US Cato Institute (he Cato Institute is a public policy research organization — think tank) developed a framework for grading US departments on their data publication practices, placing emphasis on the publication of at 'high value' datasets [17]. The main evaluation perspectives of this framework are: data availability (permanent, stable, complete, bulk accessible, incrementally accessible and open data (publicly accessible and free of proprietary encumbrances)), data authority (authoritative, timely/real-time, correctable), machine discoverability (internet-accessible, cross-referenceable) and machine processability (comprehensive conceptual data model, semantically richness, well-defined, published serializations). Solar in [33] proposes an open data maturity model to be used for assessing the commitment and capabilities of a government agency in pursuing the principles of open data; it includes three evaluation domains (each of them being divided into several sub-domains): leadership, strategy, management and legal perspective (the latter concerning external and internal regulations and licences), technological perspective (focusing on access to data, data quality and availability) and citizen-entrepreneurial perspective (degree of involvement of citizens and applications developers).

Another research stream focuses on the nature and the types of the value generated from OGD. Jetzekin [21] and [22] proposes four types of OGD value generation mechanisms, which differ in the sector generating the value (public or private), and also in the kind of generated value (economic or social): i) efficiency mechanisms (public sector organizations through OGD generate economic value by increasing internal efficiency and effectiveness), ii) transparency mechanisms (public sector organizations generate social value by offering increased transparency into government actions, which reduces 'information 'asymmetry' between government officials and citizens, and therefore misuse of public power for private benefits and corruption), iii) innovation mechanisms (private sector firms generate economic value through the creation of new products and services), iv) participation mechanisms (private sector firms generate social value through participating and collaborating with government).

However, there has been limited research towards the development of methodologies for the evaluation of OGD infrastructures/platforms, despite the big investments made by numerous government agencies all over the world. This paper contributes to filling this research gap, by developing and validating a methodology for evaluating OGD infrastructures, which adopts a 'value model' estimation approach [27] [23], takes into account the technological dimensions proposed by the abovementioned existing methodologies for evaluating OGD initiatives at the level of country or individual government agency, and combines them with the findings of the above research on the types of the value generated from OGD [21] [22].

3. Theoretical Background

For the development of our methodology we have taken into account approaches and frameworks developed from four relevant streams of previous IS research concerning: i) IS evaluation, ii) IS acceptance, iii) IS success and iv) E-services evaluation. A brief review of them is provided in this section.

Extensive research has been conducted on IS evaluation in the last twenty years [13] [16] [19] [20] [32] [38]. Its main conclusion has been that IS evaluation is a difficult and complex task, since IS offer various types of benefits, both financial and non-financial, and also tangible and intangible ones, which differ among the different types of IS. Therefore each particular type of IS requires a different evaluation methodology, which takes into account its particular capabilities and objectives. Smithson in [32] distinguishes between two basic directions of IS evaluation. The first one is 'efficiency-oriented', evaluating IS performance with respect to some predefined technical and functional specifications; it focuses on answering the question of whether the IS 'is doing things right'. The second direction is 'effectiveness-oriented', evaluating to what extent the IS provides assistance and support for the execution of business-level tasks or the achievement of business-level objectives; it focuses on answering the question of whether the IS 'is doing the right things'. The conclusions of this research stream indicate that a comprehensive methodology for evaluating a particular type of IS should include evaluation of both its efficiency and its effectiveness, based on its particular capabilities and objectives.

Another central topic in IS research has been the identification of characteristics IS that affect the intention to use them and finally the extent of their actual usage. This research has led to the development and extensive validation of the Technology Acceptance Model (TAM) and its subsequent extensions [7] [30] [36] [37][39]. According to this model the main characteristics of an IS that affect the intention to use it and finally the extent of its actual usage are: its perceived usefulness (= the degree to which users believe that using it will enhance their job performance) and its perceived ease of use (=the degree to which users believe that using it would require minimal effort). The conclusions of this IS acceptance research stream indicate that a methodology for evaluating a particular type of IS should assess its ease of use, usefulness and users' intention to use it in the future.

Another research stream that can provide useful elements is the IS success research [10] [11] [31]. The most widely used IS success model has been developed by DeLone and McLean [10]. It proposes seven IS success measures, which are structured in three layers: 'information quality', 'system quality' and 'service quality' (at the first layer), which affect 'user satisfaction' and also the 'actual use' of the IS (at the second layer); these two variables determine the 'individual impact' and the 'organizational impact' of the IS (at the third layer). Seddonin [31] proposed a re-specification and extension of this model, which includes perceived usefulness instead of actual use. The conclusions of this research stream indicate that IS evaluation should adopt a layered approach based on the above interrelated IS success measures (information quality, system quality, service quality, user satisfaction, actual use, perceived usefulness, individual impact and organizational impact) and also on the relations among them.

The emergence of numerous Internet-based e-services (e.g. information portals, e-commerce, e-banking, e-government, etc.) lead to research for the development of specialised frameworks for evaluating them [24]; extensive reviews of this research are provided by [28] and [34]. These frameworks suggest useful e-services evaluation dimensions and measures. Most of them assess the quality of the capabilities that the e-service provides to its users (being oriented towards the abovementioned efficiency evaluation). Some others assess the support it provides to users for performing various business level tasks and achieving relevant objectives (being oriented towards the abovementioned effectiveness evaluation). However, most of the above frameworks do not include advanced ways of processing the evaluation data collected from the users, in order to maximize the extraction of value related knowledge from them. They include mainly simple calculations of average values of all evaluation measures and dimensions; the relations among the proposed evaluation dimensions and measures, which could form the basis for advanced multi-dimensional statistical analysis, are not exploited at all for drawing more insights.

Only recently some research in this direction has been conducted. In [27] and [23] is proposed and verified a structured approach for assessing and improving e-services, which is based on the estimation of value models of them from users' ratings. Such a value model consists of a set of value measures, assessing the magnitude of different types of value generated by the evaluated e-service, and also of the relations among them. These value measures are organized in three layers:

(a) Efficiency layer: it includes 'efficiency' measures, which assess the quality of the basic capabilities offered by the e-service to its users.
(b) Effectiveness layer: it includes 'effectiveness' measures, which assess to what extent the e-service assists the users for completing their business-level tasks and achieving relevant objectives.
(c) Future behaviour layer: it includes measures assessing to what extent the e-service influences the future behaviour of its users (e.g. to what extent they intend to use the e-service again in the future, or recommend it to friends and colleagues).

The above value model provide a clear picture of how value generation starts through capabilities offered to the users, and then how this is transformed to support for completing their tasks and achieving their objectives, and finally how this affects their future behaviour; in this sense a value model enables a better understanding of the whole mechanism of value generation by the e-service. Also, it enables a rational definition of priorities for improvements in the capabilities it offers to users (in the first layer of the model), by giving highest priority to the improvement of those receiving lower users' ratings and at the same time having higher impact on the measures of the higher levels value (i.e. on the ones of the second and third layer). Such an approach (value models estimation) can be useful for the development of a comprehensive methodology for evaluating the OGD infrastructures, after appropriate adaptations: inclusion of quality measures of the main capabilities offered by these advanced OGD infrastructures in the first layer, and inclusion of measures of the support they provide to their users for achieving their multiple objectives in the second layer.

4. An Evaluation Methodology

A methodology for evaluating OGD infrastructures has been developed, based on one hand on the above three layers' value model approach [27] [23], and on the other hand on:

i) The approaches and frameworks from previous relevant IS research outlined in the previous section, concerning: IS evaluation (by including in the methodology both efficiency and effectiveness measures), IS acceptance (by including measures of ease of use, usefulness and future intentions), IS success (by adopting a layered evaluation approach, and including measures of both information and system quality, and also of user satisfaction and individual impact) and e-services evaluation (by including measures of both the quality of the capabilities offered to the users, and the support provided to them for achieving their OGD related objectives).

ii) The results of the analysis of potential users' requirements conducted as part of the above ENGAGE project (which, as described in more detail in [42], which include data search, provision and download capabilities, data processing capabilities, and also users – providers communication capabilities.

iii) The high level technological aspects proposed in the existing methodologies for country and government agency level OGD initiatives' evaluation (see section 2) (such as data completeness, quality, quantity, format and metadata, search capabilities, users-providers communication capabilities, users' satisfaction, platform availability).

iv) The four mechanisms of value generation from OGD proposed in [21] and [22]: efficiency, transparency, innovation and participation.

Our methodology includes the definition of a value model for these OGD infrastructures (section 4.1), and also an algorithm for estimating this value model based on users' evaluation ratings (section 4.2).

4.1. Value Model Definition

The value model consists of the main dimensions of the value that these OGD infrastructures generate, and the relations among them, organized in three value layers, adopting the structure proposed in [27] [23], which correspond to efficiency (value associated with the capabilities offered to the users), effectiveness (value associated with the support provided to users for achieving their objectives) and future behaviour (value associated with users' future behaviour) respectively; they are shown in Fig. 1.

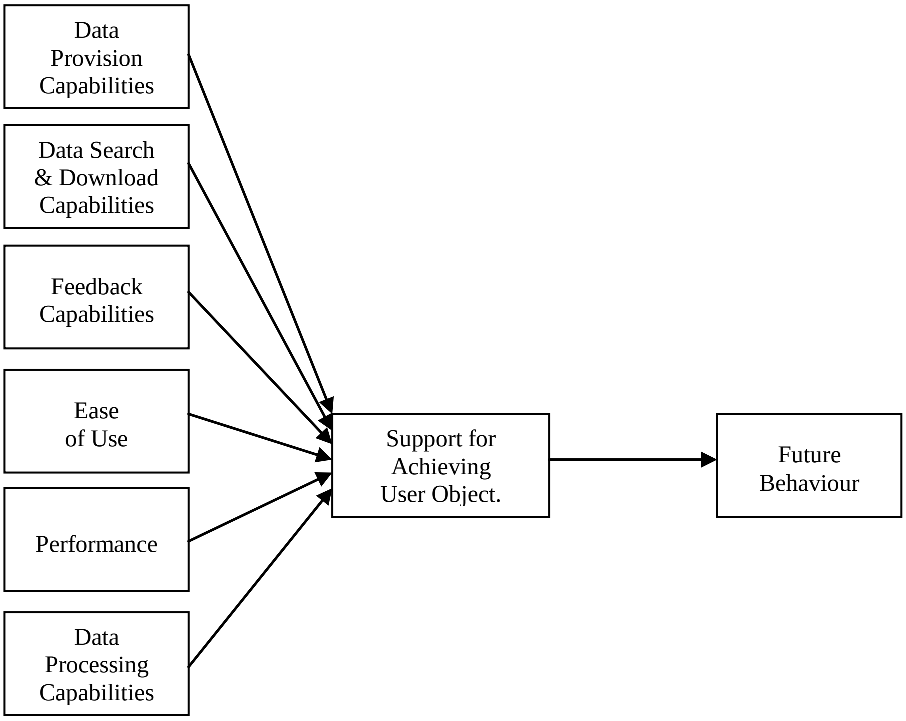

Fig. 1. Definition of value model of OGD infrastructures (main value dimensions per layer and relations among them)

The first efficiency layer includes six value dimensions in total. Four of them concern the capabilities offered by the OGD infrastructures to users: data provision capabilities (based on the 'information quality' proposed by the IS success model of DeLone and McLean [10] [11], and also in [33] [35], data search and download capabilities, data processing capabilities, feedback capabilities in [1] [42]; the other two concern ease of use (based on the TAM [7]) and performance (based on the 'system quality' proposed by the IS success model of DeLone and McLean [10] [11]). The second effectiveness layer includes one value dimension concerning the support provided by the OGD infrastructure to the users for achieving their objectives. The first layer value dimensions are expected to affect the second layer one (we can see the corresponding relations in the value model of Fig.1). Finally, the third layer includes one value dimension associated with users' future behaviour (based on the TAM [7]).

It should be noted that the value dimensions of the first efficiency layer (efficiency) are independent variables, which are under the direct control of the OGD infrastructure developer, who can take direct actions for improving them if necessary. In contrast, the value dimensions of the other two layers (effectiveness and future behaviour ones) are not under the direct control of the infrastructure developer, and are dependent to some extent on the first layer ones.

The above eight value dimensions were further elaborated, and for each of them a number of individual value measures were defined (again based on the foundations (i) to (iv) mentioned in the beginning of this section). For the second layer value dimension we defined three measures assessing the support provided for realising the second, third and fourth OGD value generation mechanisms proposed by [21][22] (see Introduction), aiming at the generation of transparency, participation and innovation related value (since the particular OGD platform developed in the ENGAGE project aimed to support mainly the generation of these three types of value, but not the development of government efficiency related value; however, we can easily add one more value measure assessing this type of value as well, in case of evaluating an OGD platform aiming to provide support in this direction). Each of these value measures was then converted to a question, which was included in a questionnaire distributed to users of the infrastructure. All these questions have the form of statements, and the users are asked to enter the extent of their agreement or disagreement with them, answering the question: “To which extend do you agree with the following statements?”. A five point Likert scale is used to measure agreement or disagreement with (i.e. positive or negative response to) such a statement (1=Strongly Disagree, 2= Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree). In Table 1 we can see the questions that correspond to the value measures of each value dimension.

It should be noted that the above value model can be adapted based on the capabilities offered by the particular OGD infrastructure developed in the ENGAGE project (e.g. additional or less value dimensions can be included according to the capabilities offered).

Table 1. Questions for Value Measures

Data Provision Capabilities (DPV)	
DPV1	The platform provides a large number of datasets
DPV2	The platform provides datasets useful to me
DPV3	The platform provides to me complete data with all required fields and detail
DPR4	The platform provides accurate and reliable data on which I can rely for my studies
DPV5	There are datasets from many different thematic areas (economy, health, education, etc.)
DPV6	There are datasets from many different countries
DPV7	The platform provides sufficiently recent data
Data Search and Download Capabilities (DSD)	
DSD1	The platform provides strong dataset search capabilities using different criteria.
DSD2	The platform provides several different categorizations of the available datasets, which assists significantly in finding the datasets I need.
DSD3	The platform enabled me to download datasets easily and efficiently.
DSD4	The datasets are in appropriate file/data formats that I can easily use.
DSD5	The datasets have also appropriate and sufficient metadata, which allowed me to understand these data and also how and for what purpose they were collected.
DSD6	The platform provides strong API for searching and downloading datasets (data and metadata)

Feedback Capabilities (FB)	
FB1	The platform provides good capabilities for giving feedback on the datasets I download, e.g. for rating datasets, for entering textual comments on them.
FB2	The platform provides good capabilities for reading available feedback of other users of datasets I am interested in, e.g. their ratings, comments.
Ease of Use (EOU)	
EOU1	The platform provides a user friendly and easy to use environment.
EOU2	It was easy to learn how to use the platform.
EOU3	The web pages look attractive.
EOU4	It is easy to perform the tasks I want in a small number of steps.
EOU5	The platform allows me to work in my own language.
EOU6	The platform supports user account creation in order to personalize views and information shown
EOU7	The platform provides high quality of documentation and online help.
Performance (PER)	
PER1	The platform is always up and available without any interruptions.
PER2	Services and pages are loaded quickly.
PER3	I did not realize any bugs while using the platform.
Data Processing Capabilities (DPR)	
DPR1	The platform provides good capabilities for data enrichment (i.e. adding new elements - fields)
DPR2	The platform provides good capabilities for data cleansing (i.e. detecting and correcting ubiquities in a dataset)
DPR3	The platform provides good capabilities for linking datasets.
DPR4	The platform provides good capabilities for visualization of datasets
Support for Achieving Users' Objectives (SUO)	
SUO1	I think that using this platform enables me to do better research/inquiry and accomplish it more quickly
SUO2	This platform allows drawing interesting conclusions on past government activity
SUO3	This platform allows creating successful added-value electronic services
Future Behaviour (FBE)	
FBE1	I would like to use this platform again.
FBE2	I'll recommend this platform colleagues.

4.2. Value Model Estimation Algorithm

The users' evaluation data collected through the above questionnaire are processed, in order to estimate the value model of the OGD infrastructure, and identify improvement priorities, using the algorithm described in this section. It consists of seven steps, which are shown in Fig. 2.

1. Initially for each value dimension we examine the internal consistency of its value measures by calculating the Cronbach Alpha of the variables corresponding to them (Boudreau 2004). This coefficient quantifies to what extent a set of variables measure different aspects of the same single uni-dimensional construct, and is calculated as:

$$\text{Alpha} = (k/(k-1)) * [1- (s^2_i)/ s^2\text{sum}] \quad (\mathbf{1})$$

where the s^2i (i = 1, 2, …, k) denote the variances of the k individual variables, while the s^2sum denotes the variance of the sum of these variables. A widely accepted and used practical 'rule of thumb' is that values of Cronbach Alpha exceeding 0.7 indicate 'acceptable' levels of internal consistency [4]. Therefore if for a value dimension its calculated value of Cronbach Alpha exceeds 0.7, we can conclude that all its measures have acceptable internal consistency; if this does not happen, we can conclude that some of the measures are not sufficiently related to this value dimension (they can be detected if for each of the individual variables is calculated the Cronbach Alpha without it, which is a standard calculation offered by all statistical packages), so they must be removed and not taken into account, or probably that this dimension should be split into two or more sub-dimensions.

2. For each value dimension an aggregate variable is calculated as the average of its individual measures' variables.

3. Average ratings are calculated for all value measures and dimensions (using for the latter the aggregate variables calculated in the previous step); this allows us to identify 'strengths' and 'weaknesses' of the OGD infrastructure (= value measures and dimensions with higher and lower average rating values respectively).

4. For each aggregate variable – value dimension of the second and third layer, we estimate a regression having it as dependent variable, and having as independent variables all the aggregate variables - value dimensions of the previous layers, in order to estimate to what extent this value dimension is affected by value dimensions of previous layers; this is quantified by the R^2 coefficient of the regression [15]. If we find that all value dimensions of the second and third layer are affected to a large extent by the value dimensions of the previous layers (e.g. having $R^2 > 0.50$), then we can conclude that this value model is characterized by coherence among its layers, so we can proceed to the following stages. On the contrary, if some value dimensions of the second or third layer are affected only to a small extent by the value dimensions of the previous layers, this indicates that some important value dimensions have been omitted in the previous layers, so we have to redefine the value model of the OGD infrastructure.

5. For each value dimension of the first level we calculate its impact on the higher level value dimensions (of the second and the third layers) using again the aggregate variables calculated in step 2.For this purpose we can use the corresponding standardised coefficients of the regressions of the above step 4. However, according to econometric literature [15], if there are high levels of correlation between the independent variables of a regression, then the estimated regression coefficients are not reliable measures of the impacts of the independent variables on the dependent variable (multi-collinearity problem). For this reason we decided to use correlations instead; so as measure of the impact of a first layer value dimension on a higher layer value

dimension has been used the correlation coefficient between them. Therefore we calculate the correlations of all first level value measures with all second and third layers' value dimensions and measure, as measures of their impact on higher level value generation.

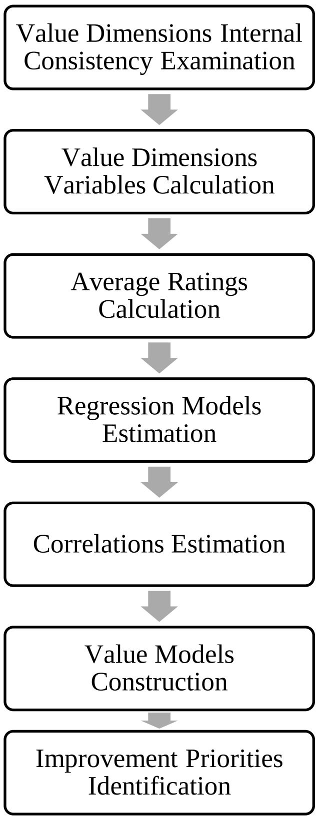

Fig.2. Value Model Estimation Algorithm

6. By combining the average ratings calculated in step 2 with the correlations calculated in step 3 we can construct a value model of the OGD infrastructure at the level of value dimensions, and also a more detailed one at the level of value measures. These models enable a deeper understanding of the value generation mechanism of the OGD infrastructure.

7. Finally the value dimensions and the value measures of the first layer, which are the only 'independent variables' within the control of the OGD infrastructure developer, are classified, initially based on their average ratings by users, and then based on their impacts on the value dimensions of the second and the third level, into four groups: low rating – high impact, low rating – low impact, high rating – high impact and high rating – low impact. The highest priority should be given to the improvement of the value dimensions and individual value measures of the first group,

which receive low ratings and at the same time have a high impact on the generation of higher layers' value; so it is on them that we should focus our scarce human and financial resources.

5. Application

The proposed methodology has been applied for the evaluation of the first version of an OGD infrastructure developed in the abovementioned ENGAGE project. The evaluation questionnaire shown in Table 1 was initially tested by three colleagues highly experienced in quantitative research in the IS domain, who found it clear and understandable, and did not report any important problems. Then 42 postgraduate students of the University of the Aegean, Greece, and the Delft University of Technology, Netherlands (both partners of the above project), all in the IS domain, were trained in the capabilities of this OGD infrastructure (in a first two hours session), and then used it for implementing an extensive scenario (in a second two hours session). Immediately after the end of these tasks they all filled the questionnaire in an online form. We believe that since all these postgraduate students had some experience in quantitative IS research, they are satisfactory sources of information concerning various aspects of value of this OGD infrastructure. The evaluation data collected from the above postgraduate students through this online questionnaire were processed according to the algorithm described above in section 4.2 using the IBM SPSS Statistics 21.0 software package.

Initially for each value dimension the Cronbach Alpha coefficient of the variables corresponding to its value measures was calculated (step 1), using the formula given in the previous section, and the results are shown in Table 2. We remark that for all value dimensions the Cronbach Alpha coefficient exceeds the minimum acceptable level of 0.7 [3]. This indicates that for all our value dimensions their value measures are sufficiently consistent, measuring different aspects of the same uni-dimensional construct. This allowed us to proceed to the calculation for each value dimension of an aggregate variable (step 2).

Table 2. Cronbach Alpha of Value Dimensions

Value Dimension	***Alpha***
Data Provision Capabilities (DPV)	0.834
Data Search and Download Capabilities (DSD)	0.805
Feedback Capabilities (FB)	0.770
Ease of Use (EOU)	0.716
Performance (PER)	0.719
Data Processing Capabilities (DPR)	0.811
Support for Achieving Users' Objectives (SUO)	0.843
Future Behaviour (FBE)	0.876

Next for all value measures and dimensions the average ratings over all respondent students were calculated (step 3), and the results are shown in the second column of Table 3 (results for value dimensions are shown in bold). We remark that according to the respondents the strongest points of this OGD platform are its ease-of use and its data processing capabilities, all perceived between moderate and good (average ratings 3.35 and 3.27 respectively). Its weakest point is its performance (with respect to its availability, response time and bugs), which is perceived as problematic (average rating 2.15). The remaining three first layer value dimensions (i.e. data provision, data search and download, and feedback capabilities) are regarded as moderate (average ratings 3.03, 3.03 and 2.97 respectively). With respect to the second layer value dimension we remark that according to the respondents this OGD platform offers between moderate and good support (closer to the former – average rating 3.17) for achieving various OGD related objectives, associated with the generation of both social value (i.e. for doing better research/inquiry and accomplishing it more quickly – average rating 3.27 – and for drawing interesting conclusions on past government activity – average rating 3.17) and economic value (i.e. for creating successful added-value electronic services – average rating 3.07).

Table 3. Average ratings of value dimensions and measures, and correlations with 2nd and 3rd layer value dimensions

Measure / Dimension	***Average ratings***	***Correl. SUO***	***Correl. FBE***	***Average Correl.***
DPV	**3.03**	**0.639**	**0.511**	**0.575**
DPV1	2.68	0.502	0.378	0.440
DPV2	3.00	0.537	0.426	0.482
DPV3	2.51	0.593	0.606	0.600
DPR4	3.02	0.544	0.375	0.460
DPV5	3.71	0.329	0.159	0.244
DPV6	3.37	0.148	0.226	0.187
DPV7	2.95	0.574	0.418	0.496
DSD	**3.03**	**0.760**	**0.747**	**0.754**
DSD1	2.68	0.516	0.520	0.518
DSD2	3.24	0.422	0.386	0.404
DSD3	3.24	0.598	0.662	0.630
DSD4	3.10	0.576	0.603	0.590
DSD5	2.90	0.589	0.549	0.569
DSD6	3.05	0.515	0.425	0.470
FB	**2.97**	**0.651**	**0.410**	**0.531**
FB1	2.90	0.622	0.284	0.453
FB2	3.05	0.624	0.442	0.533
EOU	**3.35**	**0.730**	**0.448**	**0.589**
EOU1	3.39	0.684	0.430	0.557

EOU2	3.80	0.539	0.295	0.417
EOU3	3.00	0.515	0.378	0.447
EOU4	3.39	0.487	0.293	0.390
EOU5	3.61	0.193	0.196	0.195
EOU6	3.44	0.220	0.213	0.217
EOU7	2.83	0.634	0.592	0.613
PER	**2.15**	**0.379**	**0.377**	**0.378**
PER1	2.10	0.363	0.371	0.367
PER2	2.15	0.310	0.328	0.319
PER3	2.20	0.278	0.209	0.244
DPR	**3.27**	**0.735**	**0.640**	**0.688**
DPR1	3.29	0.483	0.460	0.472
DPR2	3.26	0.644	0.581	0.613
DPR3	3.17	0.599	0.652	0.626
DPR4	3.41	0.619	0.354	0.487
SUO	**3.17**	**-**	**0.624**	
SUO1	3.27	-	0.513	
SUO2	3.17	-	0.570	
SUO3	3.07	-	0.548	
FBE	**3.19**	**0.624**	**-**	
FBE1	3.24	0.472	-	
FBE2	3.15	0.702	-	

Then we examined to what extent the value dimensions of the second and third layer are affected by the ones of the first layer (step 4). For this purpose initially we estimated one regression model having as dependent variable the second layer dimension SUO, and as independent variables the six value dimensions of the first layer. Also, we estimated one regression model having as dependent variable the value dimension of the third layer FBE and as independent variables the value dimension of the second layer, and finally another similar regression model having as additional independent variables the seven value dimensions of the first and the second layer. In Table 4 are shown the R^2 coefficients of these regression models.

Table 4. R^2 coefficients of second and third layer value dimensions' regression models

Regression Models	**R^2**
SUO model (6 indep. variables)	0.776
FBE model (1 indep. variables)	0.412
FBE model (7 indep. variables)	0.647

We can see that the R^2 coefficients of the first SUO model is 0.776, indicating that second layer value dimension is affected to a large extent by the first layer ones. On the contrary the R^2 coefficient of second FBE model has a much lower value of 0.412, indicating that the third layer value dimension is affected to a smaller extent by the one of the second layer. However, the last FBE model has a much higher R^2 coefficient 0.647, which indicates that the first and second layer value dimensions affect to a large extent the one of the third layer; therefore the first layer value dimensions affect users' future behaviour not only through the value dimensions of the second layer, but also directly as well. From the above results we can conclude that this value model is characterized by coherence among its layers.

Finally, we calculated the correlations of the first layer value dimensions and their value measures with the value dimensions of the second and third layer SUO and FBE respective (step 5), and the results are shown in the third and fourth column of Table 3 respectively. In the fifth column we can see for the first level value dimensions and measures the average of their correlations with SUO and FBE, as an indicator of their overall impact on higher level value generation. From the third column we can conclude that the data search and download capabilities, the data processing capabilities and the ease of use are the first layer value dimensions that have the strongest impact on the support provided by the OGD infrastructure to users for achieving their objectives (correlation coefficients 0.760, 0.735 and 0.730 respectively), while the performance has the weakest impact on it (correlation coefficient 0.379). From the fourth column we can conclude that the data search and download capabilities and the data processing capabilities are the first layer value dimensions that have the strongest impact on users' future behaviour (correlation coefficients 0.747 and 0.640 respectively), while the performance has the weakest impact on it (correlation coefficient 0.377). Similarly, looking at the last column of Table 3 we can conclude that the first layer value dimensions having the strongest overall impact on higher level value generation are the data search and download capabilities and the data processing capabilities (correlation coefficients 0.754 and 0.688); the performance again has the weakest impact (correlation coefficient 0.378).

Using the average ratings and correlations shown in Table 3 we construct the value model of the OGD infrastructure (step 6) at the level of value dimensions, which is shown in the Appendix (while similarly we can construct a more detailed value model at the level of value measures). It provides a compact visualization of the main dimensions/types of value generated by this e-service, their magnitudes (quantified through the corresponding average users' ratings) and the relations among them (quantified through the corresponding correlation coefficients). This enables a better understanding of the whole value generation mechanism of the OGD infrastructure, as it shows how value of one layer is transformed to value of higher layers, and in the opposite direction the origins of higher layers' value.

Furthermore, based on these average ratings and correlations of Table 3 priorities for improvements were identified (step 7). For this purpose we classified the first layer value dimensions into two groups according to their average rating: a higher ratings group and a lower ratings group (Table 5).

Also, we classified them into two groups according to their impact on (i.e. average correlation with) second and third layers' value dimensions: a higher impact group and a lower impact group (Table 6).

From Tables 5 and 6 it can be concluded that the highest improvement priority should be given to the *'data search-download capabilities'*, as they have higher impact on the generation of higher level value and at the same time receive lower ratings by users.

Table 5. Classification of first layer value dimensions according to their average ratings by the users

Lower Ratings Group	***Higher Ratings Group***
Data provision capabilities Data search-download capabilities Feedback capabilities Performance	Ease of use Data processing capabilities

Table 6. Classification of first layer value dimensions according to their impact on higher level value dimensions.

Lower Impact Group	***Higher Impact Group***
Data provision capabilities Feedback capabilities Performance	Data search-download capabilities Data processing capabilities Ease of use

6. Conclusions

Big investments are made for the development of OGD infrastructures by numerous government agencies in many countries, so it is of critical importance to proceed to comprehensive evaluation of them, in order to create a sound knowledge base about the various types of value they generate, the relations among them, and their whole value generation mechanisms in general, and also – since they are a relatively new type of IS – their improvement priorities. This knowledge will contribute to achieving higher levels of maturity in this emerging area of OGD publishing and exploitation, and finally to realizing more benefits and value from these big – and continuously growing – investments. However, as explained in more detail in the Introduction, there is limited evaluation activity in these directions, and also a lack of methodologies for this. Therefore it is quite important to develop advanced methodologies for evaluating OGD infrastructures, which provide guidance for the collection of appropriate data from their users, and also use of sophisticated quantitative techniques for processing them, in order to maximize and accelerate relevant knowledge generation in the above directions.

This paper makes the following contributions towards filling these research gaps:

i) It presents an advanced methodology for the multi-perspective evaluation of OGD infrastructures, which is quite useful, since - as mentioned above - on the one hand big investments are made by governments for the development of such infrastructures, and on the other hand there is a lack of methodologies for evaluating them and analysing the value they generate.

ii) The proposed evaluation methodology adopts a novel approach: it is based on the estimation of value models of these complex IS, which combine assessments of both the magnitudes of the main types of value that OGD infrastructures generate, and also the relations among them (which are neglected and not exploited by the 'conventional' IS evaluation approaches); this enables a deeper understanding of the whole value generation mechanism of an OGD infrastructure.

iii) It assesses both the capabilities that such an OGD infrastructure provides to its users, and also the support it offers to them for the generation of different types of both economic and social value according to previous relevant literature [21] [22].

iv) The proposed methodology not only assesses the value currently generated by an OGD infrastructure (as it usually happens with the 'conventional' IS evaluation approaches), but also provides support and direction for the increase of this value, as it enables a rational definition of improvement priorities.

Furthermore, a first application – validation of this methodology is presented, for the evaluation of an OGD infrastructure developed in the European project ENGAGE. This first application lead to interesting insights into this new type of IS, providing also evidence that our methodology can be a useful decision support tool for important ODG systems development, upgrade, improvement and management decisions. It has been concluded that strong impact on the generation of higher level value, associated with the achievement of fundamental objectives of users, and their future behaviour, have not only the 'traditional' capabilities offered by these IS (data search-download and provision capabilities), but also the 'new' ones that have recently emerged [1] [42] (data processing capabilities, and feedback capabilities associated with rating and commenting datasets and also reading other users' ratings and comments). Therefore this gradual functional enrichment of OGD infrastructures seems to be beneficial and valuable for the users. Another interesting conclusion is that such an OGD platform provides considerable support for the generation of both social and commercial value.

Our research has interesting implications for research and practice. With respect to research, it opens up a new direction of advanced future research on the multi-dimensional value generated by various types of OGD infrastructures, which enables a better understanding of its main dimensions and also the relations among them, and in general of its generation mechanisms, and provides a comprehensive framework for such research. Also, it enriches the existing body of knowledge about value models and their use for the evaluation of various types of IS. With respect to government practice, it provides a sound basis for continuous evaluation, improvement and optimization of OGD infrastructures, making optimal use of the scarce human and financial resources. Also, it allows the identification of their strengths, weaknesses and improvement priorities, and provides substantial decision support for important ODG systems related decisions. Furthermore, it provides some evidence that such OGD infrastructures can support the generation of both social and commercial value.

A limitation of our study is that the above first application - validation of the proposed OGD infrastructure evaluation methodology is based on a dataset collected from only 42 postgraduate students. Therefore further application – validation is required, based on larger and more 'professional' users' groups (more experienced than the postgraduate students' group we used in the present study), taking into account all the main segments targeted by such OGD infrastructures (e.g. professional researchers in the political, economic, administrative and management sciences, developers of added-value electronic services, political analysts and journalists). Further research is required concerning the application of the proposed methodology for the evaluation of other types of OGD infrastructures that offer different capabilities, which will necessitate adaptations of the value model definition described in 3.1.

References

1. Alexopoulos, C., Loukis, E. & Charalabidis, Y. (2014) "A Platform for Closing the Open Data Feedback Loop based on Web2.0 functionality", JeDEM 6(1): 62-68
2. Alexopoulos, C., Spiliotopoulou, L., Charalabidis, Y.,(2013),"Open data movement in greece: A case study on open governmentdata sources". In: Proceedings of the 17th PanhellenicConference on Informatics. pp. 279–286. PCI '13, ACM, New York,NY, USA, http://doi.acm.org/10.1145/2491845.2491876
3. Boudreau, M., Gefen, D. (2004), 'Validation Guidelines for IS Positivist Research', Communications of the Association for Information Systems, 13, pp. 380-427.
4. Chan, C. (2013), "From Open Data to Open Innovation Strategies: Creating e-Services Using Open Government Data", 46th Hawaii International Conference on System Sciences.
5. Commission of the European Communities (2009), "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – ICT Infrastructures for e-Science", COM (2009) 108 Final, Brussels
6. Commission of the European Communities (2011), "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Open data: An engine for innovation, growth and transparent governance", COM (2011) 882 Final, Brussels.
7. Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13(3), pp. 319-339.
8. Davies, T. (2010). "Open data, democracy and public sector reform: A look at OGD use from data.gov.uk", retrieved from http://tinyurl.com/7joks46
9. Dekkers, M., Polman, F., De Velde, R. De Vries, M. (2006), "MEPSIR – Measuring European Public Sector Information Resources: Final report of study on exploitation of public sector information – benchmarking of EU framework Conditions", Report for the European Commission, June 2006.
10. DeLone, D. H., McLean, E. R. (1992), "Information Systems Success: The Quest for the Dependent Variable". Information Systems Research, 3(1), pp. 60-95.
11. DeLone, D.H., McLean, E. R. (2003), "The DeLone and McLean model of information systems success: a ten-year update", Journal of Management Information Systems, 19, 4, 9–30.
12. Executive Office of the President (2009). Memorandum for the Heads of Executive Departments and Agencies: Open Government Directive. Retrieved from. http://www. whitehouse.gov /open/documents/open-government-directive.
13. Farbey, B., Land, F., and Targett, D. (1999), "Moving IS evaluation forward: learning themes and research issues. The Journal of Strategic Information Systems, 8(2), 189–207.
14. Fassnacht, M., and Koese, I. (2006), "Quality of electronic services", Journal of Service Research, 9(1), 19–37.
15. Greene, W. H. (2011), 'Econometric Analysis - 7th edition', Prentice Hall Inc., Upper Saddle River, New Jersey.

16. Gunasekaran, A., Ngai, E. W. T., and McGaughey (2006), “Information technology and systems justification: a review for research and applications”. European Journal of Operational Research, 173, 957–983.
17. Harper, J. (2012), ‘Grading the Government’s Data Publication Practices’, Policy Analysis – Cato Institute, No 711.
18. HM Government (2012). Open Data White Paper – Unleashing the Potential, retrieved from http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf
19. Irani, Z. (2002), “Information systems evaluation: navigating through the problem domain”. Information and Management, 40, 1, 11–24.
20. Irani, Z., and Love, P., (2008), “Information systems evaluation – a crisis of understanding” In Z. Irani and P. Love (eds.), “Evaluating Information Systems – Public and Private Sector”, Butterworth-Heinemann, UK.
21. Jetzek, T., Avital, M., Bjorn-Andersen, N. (2013a), “The Generative Mechanisms of Open Government Data”. In: Proceedings of the 21st European Conference on Information Systems (ECIS), Utrecht, The Netherlands.
22. Jetzek, T., Avital, M., Bjorn-Andersen, N. (2013b), “Generating Value from Open Government Data”. In: Proceedings of the 34th International Conference on Information Systems (ICIS), Milan, Italy.
23. Loukis, E. Pazalos, K. Salagara, A. (2012), “Transforming e-services evaluation data into business analytics using value models”, Electronic Commerce Research and Applications, 11(2), 129–141.
24. Lu, J., and Zhang, G. (2003), “Cost benefit factor analysis in e-services”, International Journal of Industry Service Management, 14, 5, 570–595
25. Nentwich, M (2003), “Cyberscience: Research in the Age of the Internet”, Vienna, Austrian Academy of Sciences.
26. Open Knowledge Foundation (OKF) (2012). “Open Data Handbook version 1.0.0” (revised November 14, 2012). Available at http://opendatahandbook.org/pdf/OpenDataHandbook.pdf
27. Pazalos, K., Loukis, E., Nikolopoulos, V. (2012), ‘A Structured Methodology for Assessing and Improving e-Services in Digital Cities’, Telematics and Informatics Vol. 29, pp. 123-136.
28. Rowley, J. (2006), ‘An analysis of the e-service literature: towards a research agenda’. Internet Research, 16, 3, 2006, 339–359.
29. Saha, R., Grover, S. (2011), “Quantitative Evaluation of Website Quality Dimension for Web 2.0”, International Journal of u- and e- Service, Science and Technology, 4(4), 15-36.
30. Schepers, J. and Wetzels, M. (2007). ‘A meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects’. Information & Management, 44, pp. 90-103.
31. Seddon, P. B. (1997), “A respecification and extension of the DeLone and McLean model of IS success”. Information Systems Research, 8(3), 240–253.
32. Smithson, S., and Hirscheim, R. (1998), “Analysing information systems evaluation: another look at an old problem. European Journal of Information Systems, 7, 158–174.
33. Solar, M., Concha, G., Meijueiro, L. (2013), ‘A Model to Assess Open Government Data in Public Agencies’, in H. J. Scholl et al. (Eds.) Proceedings of IFIP EGOV Conference 2012, LNCS 7443, pp. 210-221.
34. Sumak, B., Polancic, G., and Hericko, M. (2009), “Towards an e-service knowledge system for improving the quality and adoption of e-services”, In Proceedings of the 22nd Bled ‘eEnablement: Facilitating an Open, Effective and Representative Society’, June 14–17, 2009, Bled, Slovenia.
35. Ubaldi, B. (2013), “Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives”. OECD Working Papers on Public Governance No 22.
36. Venkatesh, V., and Davis, F. D. (2000), “A theoretical extension of the technology acceptance model: four longitudinal field studies. Management Science, 45(2), 186–204.

37. Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003) "User acceptance of information technology: toward a unified view", MIS Quarterly, 27, 3, 2003, 425–478.
38. Willcocks, L., and Graeser, V. (2001), "Delivering IT and E-business Value". Butterworth–Heinemann, Boston, MA.
39. Wixom, B. H., Todd, P.A. (2005), "A Theoretical Integration of User Satisfaction and Technology Acceptance", Information Systems Research, 16(1), pp. 85-102.
40. World Bank (2013), "Open Data Readiness Assessment Methodology", retrieved from http://data.worldbank.org/sites/ default/files/1/od_readiness_-_revised_v2.pdf.
41. World Wide Web Foundation (2015), "Open Data Barometer Global Report – Second Edition", retrieved from http://www. opendataresearch.org/barometer.
42. Zuiderwijk, A., Janssen, M., Jeffery, K., (2013), "An e-infrastructure to support the provision and use of open data", International Conference for eDemocracy and Open Government 2013 (CEDEM 13), 22-24 May 2013, Krems, Austria.

Charalampos Alexopoulos is PhD Candidate at University of the Aegean and Research Associate in the Information Systems Laboratory (ISL/AEGEAN-RU), working on high-level policy making, research and pilot application FP7 and national projects (ENGAGE, SHARE-PSI 2.0, EU-COMMUNITY, PADGETS, NOMAD, NET-EUCEN, PLUG-IN). He is a computer science graduate from the University of Peloponnese with MSc in management information systems from University of Aegean and he has published in scientific conferences and journals on IS evaluation, open data, semantic interoperability and e-Government. He has also been a member of the organising, reviewing and scientific committees of several conferences (ICTDM, Samos Summit, IEEE SOLI, SSCORE, ECIS).

Euripidis Loukis is Associate Professor of Information Systems and Decision Support Systems at the University of the Aegean. Formerly he has been Information Systems Advisor at the Ministry to the Presidency of the Government of Greece, Technical Director of the Program of Modernization of Greek Public Administration of the Second Community Support Framework and National Representative of Greece in the programs 'Telematics' and 'IDA' (Interchange of Data between Administrations) of the European Union. He is the author of numerous scientific articles in international journals and conferences. His current research interests include e-government, e-participation, information systems impact and internal/external determinants, business process adaptation and medical decision support systems.

Yannis Charalabidis is Assistant Professor in the Department of Information and Communication Systems Engineering of the University of Aegean. In parallel, he serves as Director of the Innovation and Entrepreneurship Unit of the University, designing and managing youth entrepreneurship activities, and Head of Information Systems Laboratory, coordinating policy making, research and pilot application projects for governments and enterprises worldwide. He has more than 20 years of experience in designing, implementing, managing and applying complex information systems as project manager, in Greece and Europe. He has published more than 150 papers in international journals and conferences, while actively participating in international standardisation committees and scientific bodies.

Received: January 10, 2015; Accepted: December 10, 2015.

Appendix

Estimated value model of the evaluated OGD infrastructure