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

Euripidis Loukis
University of Aegean, Department of Information and Communication Systems Engineering
Gorgiras Str., Liberis Building, Karlovasi, Samos, GR-83200

Yannis Charalabidis
University of Aegean, Department of Information and Communication Systems Engineering
Gorgiras Str., Liberis Building, Karlovasi, Samos, GR-83200

Charalampos Alexopoulos
University of Aegean, Department of Information and Communication Systems Engineering
Gorgiras Str., Liberis Building, Karlovasi, Samos, GR-83200

Abstract

Many government agencies worldwide have started making considerable investments for developing information systems that enable opening important data they possess to the society, in order to be used for scientific, commercial and political purposes. In order to rationalise and support future decisions concerning the development, upgrade, improvement and management of this new type of information systems it is important to understand better what value they create and how, and at the same time to identify the main improvements they require. This paper contributes in this direction 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. It consists of several value dimensions and their corresponding value measures, organized in three 'value layers', and also the relations among them. These three value layers concern value related to the efficiency of the OGD (= quality of the various capabilities it provides to the users), its effectiveness (= degree of supporting users for achieving their objectives) and also users' future behavior intentions respectively. The proposed methodology has been applied successfully to an advanced OGD system developed as part of the European project ENGAGE ('An Infrastructure for Open, Linked Governmental Data Provision towards Research Communities and Citizens'), providing to interesting insights and improvement priorities. This first application provides evidence that our methodology can be a useful decision support tool for important ODG systems development, upgrade, improvement and management decisions.

KEYWORDS

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

1. INTRODUCTION

Many government agencies worldwide have started making considerable investments for developing information systems that enable opening important data they possess to the society, in order to be used for scientific, commercial and political purposes [1]-[6]. The future evolutions in this area will rely critically on the decisions of many government agencies concerning the development of new open government data (OGD) systems, the upgrade (e.g. of initial small pilot ones, so that they can serve more users and host more datasets), the functional improvement (e.g. in order to serve better users' needs) and the management (e.g. concerning the level of maintenance and technical support) of existing ones. In order to rationalise and support these future decisions it is important to understand and assess the various types of value that these OGD systems generate, their value generation mechanism, 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 [1] 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. The same study, in order to contribute to filling this gap, proposes an analytical framework for assessing country-level OGD initiatives, which includes three main assessment dimensions: strategy and legal-institutional framework, implementation framework, and value creation (social, political and economic). Also, in [7] is described 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 consisting of several individual variables): establishment and legal perspective, technological perspective and citizen-entrepreneurial perspective. Therefore, though there are some first methodologies for evaluating OGD initiatives at the level of country and individual government agency, there is not a methodology for evaluating OGD systems, which is the most critical level for value creation from OGD.

In this direction this paper describes and validates an methodology for evaluating OGD infrastructures, which adopts the 'value model' approach to IS evaluation proposed in [10-11]. 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; this leads to the formation of a value model of the IS, which provides highly important advantages: it enables a deeper understanding of the whole value generation mechanism of the particular IS, and also a rational definition of its improvement priorities (see section 2 for more details on this approach). The proposed methodology has been used for the evaluation of an advanced 'second generation' OGD e-Infrastructure [8-9] developed in the European project ENGAGE (for more details see http://www. engagedata.eu/about/).

In the following section 2 the proposed methodology is described, while in section 3 the above application of it is presented. Finally in section 4 the conclusions are summarized and future research directions are proposed.

2. AN EVALUATION METHODOLOGY

Our methodology for evaluating this advanced second generation of OGD infrastructures was based on one hand on the above three layers' value model approach [10-11], and on the other hand on:

i) Approaches and frameworks from previous relevant IS research (briefly reviewed in [10-11]) concerning: IS evaluation (we have included in the methodology both IS 'efficiency' and 'effectiveness' measures), IS acceptance (we have included measures of ease of use, usefulness and users' future intentions), IS success (our methodology has adopted a layered evaluation approach, and has included measures of both information and system quality, and also of user satisfaction and individual impact) and e-services evaluation (our methodology has included 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 [8-9], include data search, provision and download capabilities, data processing capabilities, and also capabilities for communication with other users).

iii) The high level technological aspects proposed in the methodologies for country and government agency level OGD initiatives' evaluation proposed in [1] and [7] respectively (such as data completeness, quality, quantity, format and metadata, search capabilities, users satisfaction, platform availability).

2.1 Value Model Definition

Based on the above foundations, our methodology includes the definition of a value model of an advanced second generation OGD infrastructures, which consists of the main dimensions of the value it generates, and the relations among them, and is shown in the Appendix. We remark that these value dimensions are organized in three value layers, adopting the structure proposed by [10-11], which correspond to efficiency (value associated with the capabilities it offers to the users), effectiveness (value associated with the support provided to the users for achieving their objectives) and future behavior (value associated with users' future behavior) respectively. It should be noted that the value dimensions of the first efficiency layer 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 are not under the direct control of the infrastructure developer, and are dependent to some extent on the first level ones.

The presented 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 iii) mentioned in the beginning of this section). Each of these value measures was then converted to a question to be included in a questionnaire to be 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 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 mentioned that the above value model definition can be adapted based on the capabilities offered by the particular OGD infrastructure under evaluation, so additional value dimensions can be added for additional capabilities that might be offered (or some value dimensions might not be used if the corresponding capabilities are not offered).

2.2 Value Model Estimation Algorithm

Additionally, an algorithm has been developed for estimating this value model based on users' evaluation ratings, adopting the approach proposed in [10-11]. In particular, these users' evaluation data collected through the above questionnaire will be 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 as follows:

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 its value measures [12]. This coefficient quantifies to what extent a set of variables measure different aspects of the same single uni-dimensional construct, and is calculated as:

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

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 of the variables [12]. 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.

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' (=value measures and dimensions with high and low average ratings) of the OGD infrastructure.

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 [13]. 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 (e.g. R^2 <0.5), 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 standardized coefficients of the regressions of the above step 4. However, according to the econometric literature [13], if there are high levels of correlation between the independent variables of a regression (and this happened in our data), 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 one has been used the correlation coefficient between them. Furthermore we calculated the average correlations of each first level value measures with all second and third layers' value dimensions and measure, as measures of their impact on higher level value generation.

6. By combining the average ratings calculated in step 2 with the correlations calculated in step 3 we can construct one 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 value models enable a deeper understanding of the value generation

mechanism of the OGD infrastructure, as they visualize on one hand the levels of the various types of value it generates, and on the other hand how value of one layer is transformed to value of higher layers, and also the origins of higher layers' value in the lower layers.

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, 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 level value; so it is on them that we should focus our scarce human and financial resources. Furthermore, for each value dimension and value measure we can calculate an 'Improvement Priority Index' ($IMPR_i$), which quantifies the priority that has to be given to it, and is equal to the product of its average correlation with second and third layer value dimensions ($AVCOR_i$) and the distance of its average rating ($AVRAT_i$) from the highest possible rating (equal to 5 in our case):

$$IMPR_i = AVCOR_i * (5 – AVRAT_i)$$

3. APPLICATION

The proposed methodology has been applied for the evaluation of the first version of an advanced second generation 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 (The Netherlands) (both partners of the above project) in the IS domain were trained in the capabilities of this OGD infrastructure (in a two hours session), and used it next for implementing a representative scenario (in an one hour session). Immediately after the end of these tasks they all filled the questionnaire in paper 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 collected evaluation data were processed using the algorithm described in section 2.2. Initially we calculated the Cronbach Alpha values for all dimensions (step 1), and since all of them exceeded the lowest acceptable value 0.7 we can conclude that they are internally consistent, so we proceeded to the calculation of the value dimensions' aggregate variables (step 2). Then we estimated the regression models of the second and the third layer value dimensions (step 4), and both had R^2 coefficients higher than 0.5. In Table 1 we can see for each value dimension and each value measure its average rating (step 3), its correlations with the two second and third layer measures and also the average of these two correlations (step 5).

Table 1. Questions-Average ratings of value dimensions/measures, and correlations with 2nd and 3rd layer value dimensions

Dimension / Measure		Average ratings	Correl. SUO	Correl. FBE	Average Correl.
Data Provision Capabilities (DPV)		**3.03**	**0.639**	**0.511**	**0.575**
DPV1	The platform provides a large number of datasets	2.68	0.502	0.378	0.440
DPV2	The platform provides datasets useful to me	3.00	0.537	0.426	0.482
DPV3	The platform provides to me complete data with all required fields and detail	2.51	0.593	0.606	0.600
DPR4	The platform provides accurate and reliable data on which I can rely for my studies	3.02	0.544	0.375	0.460
DPV5	There are datasets from many different thematic areas (economy, health, education, etc.)	3.71	0.329	0.159	0.244
DPV6	There are datasets from many different countries	3.37	0.148	0.226	0.187
DPV7	The platform provides sufficiently recent data	2.95	0.574	0.418	0.496
Data Search and Download Capabilities (DSD)		**3.03**	**0.760**	**0.747**	**0.754**
DSD1	The platform provides strong dataset search capabilities using different criteria.	2.68	0.516	0.520	0.518

DSD2	The platform provides several different categorizations of the available datasets, which assists significantly in finding the datasets I need.	3.24	0.422	0.386	0.404
DSD3	The platform enabled me to download datasets easily and efficiently.	3.24	0.598	0.662	0.630
DSD4	The datasets are in appropriate file/data formats that I can easily use.	3.10	0.576	0.603	0.590
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.	2.90	0.589	0.549	0.569
DSD6	The platform provides strong API for searching and downloading datasets (data and metadata)	3.05	0.515	0.425	0.470
User-level Feedback Capabilities (UFB)		**2.97**	**0.651**	**0.410**	**0.531**
UFB1	The platform provides good capabilities for giving feedback on the datasets I download, e.g. for rating datasets, for entering textual comments on them.	2.90	0.622	0.284	0.453
UFB2	The platform provides good capabilities for reading available feedback of other users of datasets I am interested in, e.g. ratings, comments.	3.05	0.624	0.442	0.533
Ease of Use (EOU)		**3.35**	**0.730**	**0.448**	**0.589**
EOU1	The platform provides a user friendly and easy to use environment.	3.39	0.684	0.430	0.557
EOU2	It was easy to learn how to use the platform.	3.80	0.539	0.295	0.417
EOU3	The web pages look attractive.	3.00	0.515	0.378	0.447
EOU4	It is easy to perform the tasks I want in a small number of steps.	3.39	0.487	0.293	0.390
EOU5	The platform allows me to work in my own language.	3.61	0.193	0.196	0.195
EOU6	The platform supports user account creation in order to personalize views and information shown	3.44	0.220	0.213	0.217
EOU7	The platform provides high quality of documentation and online help.	2.83	0.634	0.592	0.613
Performance (PER)		**2.15**	**0.379**	**0.377**	**0.378**
PER1	The platform is always up and available without any interruptions.	2.10	0.363	0.371	0.367
PER2	Services and pages are loaded quickly.	2.15	0.310	0.328	0.319
PER3	I did not realize any bugs while using the platform.	2.20	0.278	0.209	0.244
Data Processing Capabilities (DPR)		**3.27**	**0.735**	**0.640**	**0.688**
DPR1	The platform provides good capabilities for data enrichment (i.e. adding new elements - fields)	3.29	0.483	0.460	0.472
DPR2	The platform provides good capabilities for data cleansing (i.e. detecting and correcting ubiquities in a dataset)	3.26	0.644	0.581	0.613
DPR3	The platform provides good capabilities for linking datasets.	3.17	0.599	0.652	0.626
DPR4	The platform provides good capabilities for visualization of datasets	3.41	0.619	0.354	0.487
Support for Achieving User-level Objectives (SUO)		**3.17**	**-**	**0.624**	
SUO1	I think that using this platform enables me to do better research/inquiry and accomplish it more quickly	3.27	-	0.513	
SUO2	This platform allows drawing interesting conclusions on past government activity	3.17	-	0.570	
SUO3	This platform allows creating successful added-value electronic services	3.07	-	0.548	
Future Behaviour (FBE)		**3.19**	**0.624**	-	
FBE1	I would like to use this platform again.	3.24	0.472	-	
FBE2	I'll recommend this platform colleagues.	3.15	0.702	-	

Using the average ratings and correlations shown in Table 1, we constructed 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). This provides a compact visualization of the value generation mechanism of this OGD infrastructure.

Furthermore, 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 2). Also, we classified them into two groups according to their impact on (average correlation with) second and third layers' value dimensions: a higher impact group and a lower impact group (Table 3). From these two classifications we can conclude that our highest priority should be given to the improvement of the data search-download capabilities, since they received low ratings from the users, and at the same time they have high impact on higher layers' value generation.

Table 2. 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 capabil. User-level feedback capabilities Performance	Ease of use Data processing capabilities

Table 3. Classification of first layer value dimensions according to their impact on higher level value dimensions

Lower Impact Group	Higher Impact Group
Data provision capabilities User-level feedback capabilities Performance	Data search-download capabil. Data processing capabilities Ease of use

Finally, we calculated the Improvement Priority Index for all value dimensions, shown in Table 4 in order of improvement priority; so the highest improvement priority should be given to the data search-download capabilities (confirming the conclusion drawn based on Tablew 3 and 4), mainly due to its high impact on higher layer value generation, followed by data processing and data provision capabilities.

Table 4. Improvement Priority Index for all value dimensions

Value Dimension	Improvement Priority Index
Data search-download capabilities (DSD)	1.49
Data processing capabilities (DPR)	1.19
Data provision capabilities (DPV)	1.13
User-level feedback capabilities (UFB)	1.08
Performance (PER)	1.08
Ease of Use (EOU)	0.97

4. CONCLUSIONS

In the previous sections of this paper has been presented and validated an OGD evaluation methodology that adopts a novel approach, based on the estimation of value models of these advanced OGD infrastructures, which include assessments of both the main types of value they generate, and also the relations among them (which are neglected and not exploited by the 'conventional' IS evaluation approaches). It enables not only the identification of strengths and weaknesses of an OGD infrastructure, but also a deeper understanding of the whole value generation mechanism of it. It also allows a rational definition of improvement priorities, which is quite important, as this is a relatively new type of IS.

The first application of the proposed evaluation methodology for the evaluation of the users' perspective of an advanced second generation OGD Infrastructure, lead to interesting insights into this new type of IS, especially with respect to their novel features, providing evidence that our methodology can be a useful decision support tool for important ODG systems development, upgrade, improvement and management decisions. In particular, it has been concluded that the data processing capabilities, a key novel feature of this new generation of OGD Infrastructures, has a very strong impact on the generation of higher level value, associated with the achievement of fundamental objectives of users, and their future behaviour. Another novel feature, the user-level feedback capabilities (concerning rating and commenting datasets and also reading other users' ratings and comments), was found to have considerable impact on higher level value generation.

Further research is required concerning the application of the proposed methodology for the evaluation of the other advanced second generation OGD infrastructures, after appropriate adaptations. Furthermore, the above future research should be 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).

5. REFERENCES

[1] Ubaldi, B. (2013), "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives". OECD Working Papers on Public Governance No 22.

[2] 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.

[3] 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.

[4] 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

[5] Nentwich, M (2003), "Cyberscience: Research in the Age of the Internet", Vienna, Austrian Academy of Sciences.

[6] Chan, C. (20130, "From Open Data to Open Innovation Strategies: Creating e-Services Using Open Government Data", 46^{th} Hawaii International Conference on System Sciences.

[7] 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.

[8] Charalabidis, Y., Ntanos, E., Lampathaki, F. (2011). An architectural framework for open governmental data for researchers and citizens. In M. Janssen, A. Macintosh, J. Scholl, E. Tambouris, M. Wimmer, H. d. Bruijn & Y. H. Tan (Eds.), Electronic government and electronic participation joint proceedings of ongoing research and projects of IFIP EGOV and ePart 2011, pp. 77-85.

[9] Zuiderwick, 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.

[10] 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.

[11] 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.

[12] Boudreau, M., Gefen, D. (2004), 'Validation Guidelines for IS Positivist Research', Communications of the Association for Information Systems, 13, pp. 380-427.

[13] Greene, W. H. (2011), 'Econometric Analysis - 7th edition', Prentice Hall Inc., Upper Saddle River, New Jersey.

Appendix

Value Model of an Advanced Second Generation OGD Infrastructure