Decision-making Support for Opening Government Data

Luthfi, A.

DOI
10.4233/uuid:48e47c01-34f3-41e3-88e3-1180aecfec4b

Publication date
2021

Document Version
Final published version

Citation (APA)
Luthfi, A. (2021). *Decision-making Support for Opening Government Data*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:48e47c01-34f3-41e3-88e3-1180aecfec4b

DECISION-MAKING SUPPORT FOR OPENING GOVERNMENT DATA

AHMAD LUTHFI

Decision-making Support for Opening Government Data

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Wednesday 22 September 2021 at 17.30

by
Ahmad LUTHFI

Master of Computer Science, Gadjah Mada University, Indonesia
born in Curup, Indonesia

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairman
Prof.dr.ir. M.F.W.H.A. Janssen	Delft University of Technology, promotor
Prof.dr.ir. J.W.H. Crompvoets	KU Leuven, promotor

Independent members:

Dr. D. Witarsyah Jacob	ODRC Telkom University, Indonesia
Prof.dr. H. Lee	Brunel University London, UK
Prof.dr. W.M. de Jong	Erasmus University Rotterdam
Prof.dr. Y. Tan	Delft University of Technology
Prof.dr.ir. P.H.A.J.M. van Gelder	Delft University of Technology

Keywords: decision-making support, open data, open government data, stakeholders, advantages, disadvantages, costs, benefits, risks, Bayesian-belief networks, fuzzy multi-criteria decision making, decision tree analysis

This research was funded by Indonesia Endowment Fund for Education in collaborating with the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia (LPDP–BUDI LN).

ISBN [978-94-6384-244-0]
Printed by IPSKAMP Printing (https://proefschriften.net/)
An electronic version of this dissertation is available at https://repository.tudelft.nl/

Summary

Government institutions collect and produce an extraordinary number of datasets to conduct and execute their programs and agendas. Various types of datasets collected by the governments can increase transparency and accountability, improve citizen engagement, and create value-added services for the public. Through the Open Government Data (OGD) initiatives, Non-Government Organisations (NGOs), private agencies, business enablers, data analysts, researchers, civil societies, and other open data stakeholders can take advantage of disclosing the government datasets.

Despite its significance, the decision-making process to disclose government datasets is given limited attention and encounters several challenges. Although numerous datasets have been published to the public, many datasets remain undisclosed. Government institutions face several challenges in deciding to open datasets. First, the governments have not systematically analysed datasets to identify the benefits and disadvantages of opening datasets. Decision-makers, policy-makers, civil servants, and administrative officers do not know how to balance the advantages and disadvantages of opening datasets. Second, various stakeholders' backgrounds may have different objectives and interests to analyse and disclose datasets. Third, the easy understanding of possible disadvantages of opening datasets results in moving away from the potential benefits due to the risk-avoiding culture in government. Therefore, these results in keeping datasets undisclosed.

Furthermore, the stakeholders' involvement in the decision-making process to open data, such as politicians, executive boards, decision-makers, civil servants, data analysts, and societies, all play essential roles and have different objectives for opening and using the datasets. For example, some decision-makers might have the authority to publish or keep the dataset closed. Some public servants might be risk-averse, whereas others might open datasets without considering possible negative consequences. As a result, the decision-making process becomes fuzzy, and the objectives of disclosing data are not reached. The different roles and interests of the

heterogonous actors in the internal government organisation might create uncertainty and delay the decision-making process.

Although there are guidelines, there are no decision-making tools to help governments decide to open their datasets. On the governments' side, the potential disadvantages might easily dominate over the advantages. It is much easier for the decision-makers to keep a dataset closed than take the disadvantages of releasing a dataset. The lack of insights and expertise in estimating the potential advantages and disadvantages of opening data can also lead to uncertainty, which might result in avoiding the disclosure of datasets. Therefore, this research aims to develop Decision-making Support for Opening Government Data (DSOD). This DSOD accommodates a systematic approach to decide to open datasets. To achieve the objective of this research, we followed the Design Science Research (DSR) approach. The DSR approach results in developing a prototype of the DSOD as a design artefact and demonstrate it to the stakeholders.

This study successfully answered four main research questions, as follows:

First, *what are the advantages and disadvantages of opening data?* A comprehensive systematic literature review was performed to answer the first research question (RQ#1). Various types of open data's advantages, such as improving transparency, enhancing accountability, and stimulating citizen engagement, were identified. We also identified the disclosing data's disadvantages, such as privacy violation, data-sharing dispute, and misinterpretation of the data. These identified influencing factors contribute to the literature, particularly for developing a taxonomy of the advantages and disadvantages of opening data.

Second, *what are the elements of the decision-making support for opening data?* A comprehensive systematic literature review and a preliminary case study in some government departments in Indonesia were carried out to answer the second research question (RQ#2). This question's answering resulted in identifying eight main elements from the literature, namely (1) Database Management System (DBMS), (2) Model Base Management system (MBMS), (3) Dialogue Generation and Management System (DGMS), (4) User interface, (5) User authentication, (6) Decision context, (7), Knowledge-based, and (8) The model and analytical tool. The DSOD's elements were used to formulate detailed steps that can be followed for making decisions to open data. The best practice on how decision-makers in Indonesia decide to open data was used as a starting point to define the decision-support elements. The elements of the DSOD contribute to the literature with regards to the decision-making support requirements and detailed steps that need to be taken in the decision-making process to open data.

Third, *what are the functionalities of a prototype?* The design of a DSOD was implemented using a prototype. Several functionalities of the DSOD were provided to answer the third research question (RQ#3). The proposed prototype covered a conceptual model of decision-making to open data, which consist of five main functionalities, namely (1) retrieve and decompose datasets, (2) evaluation the datasets, (3) assessment the datasets, (4) decision-making, and (5) iteration and update the datasets. The functionalities of the DSOD contribute to the literature and practical overview by employing three decision-making methods, namely Bayesian-belief Networks (BbN), Fuzzy Multi-criteria Decision-making (FMCDM), and Decision Tree Analysis (DTA).

Fourth, *what are the differences between BbN, FMCDM, and DTA to support decision-making about opening of the dataset?* To answer RQ#4, we experimented with three different methods using three stakeholder groups. The stakeholder groups are governments, academia/universities, and communities/professionals. There are three main criteria for evaluating the differences between the three selected methods, e.g., how transparent is the process? How accurate are the expected results? And how useful is the proposed DSOD is for the open government data stakeholders?

The quasi-experiment indicated that stakeholders who had limited knowledge found the advanced decision-making methods challenging to use. These types of stakeholders, such as civil servants and administrative officers, prefer to employ the DOSD for its usefulness and easiness. In contrast, stakeholders having already knowledge and expertise in the decision-making process, focused more on the transparency and accuracy of decisions. Our findings suggest that the preferred problem-solving strategies depend on the stakeholder’s background and expertise.

The use of three different methods was evaluated. A quasi-experiment was conducted by applying a two-group random assignment pre-test and post-test design. The outcomes confirm that there are differences in performance. BbN can weigh the benefits and disadvantages of the opening dataset by taking into account uncertainties and conditional dependencies, which are driven by external events, such as the opening of other datasets. Our quasi-experiment shows that BbN is the best method for understanding how the factors are related to each other. FMCDM provides an efficient solution to decision-making problems in the open government data field that can work with various types of inputs, such as qualitative and quantitative information from the experts. These results in benefits for the stakeholders that have different knowledge levels and expertise. The DTA provides a mechanism for dealing with the costs, benefits, and possible consequences of opening data. This results in stakeholder benefits to create a comprehensive analysis of the implications of the potential costs and benefits of opening data along each decision branch and identifies decision nodes. The three methods are the most appropriate depending on the goals to achieve, the stakeholders who want to use the method, the datasets, and the situation. The findings contribute to a better understanding of the effects of decision-making methods.

In the e-procurement case study, our quasi-experiment identified several essential findings. First, in terms of the transparency process of the DSOD, BbN appears to be the best method to be employed in the decision-making process to analyse the datasets compare to two other methods. Second, related to the accuracy, BbN presented the best performance to guarantee the process's results are more precise. Third, regarding the perceived ease of use, the BbN seems the best method. Fourth, however, the differences in the tendency to use the usefulness methods, DTA appears to be the best systematic method.

Moreover, in the medical records case study, transparency process of the DSOD, DTA performs to be the best method. In terms of the accuracy of the results, BbN scored highest compared to other methods. Regarding the perceived ease of use, the DTA was successfully performed in the DSOD experiment. Finally, BbN has performed the best method to analyse the advantages and disadvantages of opening data with regard to its usefulness.

Our MANOVA analysis shows that in the case study of e-Procurement, all the independent variables (transparency, accuracy, perceived ease of use, and usefulness) have no significant difference in the three methods used (BbN, FMCDM, and DTA). The MANOVA analysis indicates that there were no significant differences among the methods used for the DSOD. On the contrary, in the second case study (medical records dataset), the MANOVA analysis shows that three independent variables, e.g. accuracy, perceived ease of use, and usefulness, significantly differ regarding the three methods used (BbN, FMCDM, and DTA). At the same time, the transparency variable has no significant differences among the proposed methods. The methods perform differently on the level of accuracy, perceived ease of use and usefulness.

Our findings contribute to theorising decision-making for opening data. First, the taxonomy of OGD provides a comprehensive overview of the benefits and disadvantages. Second, the taxonomy can be used as the basis for decision-making support by systematically taken the benefits and disadvantages into account. Third, systematic and detailed steps of decision-making support to open data developed in this study can be used by the government organisations to analyse the diverse datasets. This provides a more fine-grained decision than merely closing or opening datasets. Fourth, the comparison of BbN, FMCDM and DTA reveals their different impacts and advantages for decision-makers.

There are three limitations in this study that can be addressed in future research. First, this study was carried out from the interpretive paradigm, which supports multiple insights and realities derived from human beings. The interpretivism approach in this research is essential to understand the decision-making process's circumstances to open data. However, interpretative research has been criticised for not evaluating the objective evaluation criteria. Therefore, the different stakeholders' subjectivity in determining the disadvantages and advantages elements has openly occurred.

Second, we focussed on decision-making support, but we did not analyse the political process resulting in open data policies in the decision-making process. Accordingly, power, a sense of openness, and dominance can play a role in the decision-making process.

Third, we focussed on public servants in Indonesia. In other countries, the decision-making process might be different, and persons' traits might be different. In some cultures, risk-averseness might be dominated, whereas in other countries, transparency. Therefore, bias might have occurred in this research because of the demographics scope of experimental-case study participants.

Our first future research recommendation is to use the DSOD in other contexts and settings to generalise the findings. Second, there were only a few prior research studies in decision-making support to open data. As a result, further research is recommended to develop a research typology of participant groups. The use of research typology makes more specific the different types of participant groups and their requirements. Regarding the focused methods used in this research, other methods might be relevant, or methods can be combined to arrive at even better decision support. Hence, we recommend further research in this area. Finally, the different outcomes in the two experiments result in the recommendation to conduct more experiments in different contexts.

Preface and acknowledgements

My research would have been impossible without the support, help, and guidance of several individuals and institutions who contributed their valuable assistance in preparing and completing this dissertation.

First and foremost, I would like to express my sincere gratitude to Marijn Janssen for giving me a wonderful opportunity to complete my PhD dissertation under your supervision. Thank you for all the ideas, knowledge, expertise, and patience in supervising me to complete this dissertation. Your insightful feedback encouraged me to sharpen my academic thinking and brought my work to a higher level. You supervised me on how to become an independent researcher. It was an honour that you promoted me to the broader research groups through IDEA League doctoral school and open data conferences. Your kind attention to my qualifications and my family makes me feel at home in the Netherlands.

I would also like to express my heartiest gratitude to my co-promotor Joep Crompvoets. Thank you for keeping me highly motivated to complete my PhD dissertation. Joep, I am sincerely grateful for your presence and continuing support during my PhD journey. Your positive outlook and confidence in my research inspired me to keep progress with my dissertation in a convenient environment.

Further, I would like to extend my thanks to my colleagues of the ICT Section of the Department of Engineering Systems and Services, Faculty of Technology, Policy and Management of the Delft University of Technology. I am deeply grateful to Yao-Hua Tan and Boriana Rukanova for giving me the opportunity to collaborate on writing a paper together and evaluating my decision-making support tool at Unit Data Analytics and Big Data, Customs National Office, the Netherlands. I also owe outstanding gratitude to Mark de Reuver, Anneke Zuiderwijk, and Jolien Ubacht for their constructive feedback and friendly discussion during my research presentations in the ICT Colloquium several times.

I would also like to extend my gratitude to other ICT research group partners. Just to mention few Darusalam, Agung Wahyudi, Rizal Batubara, Yi-Yin, Arie

Purwanto, Dhata Praditya, Fernando Kleiman, Mohammad Ali Latifi, Clara Maathuis, Ricardo Matheus, Sélinde Engelenburg, Ade Ratnasari, Wirawan Agahari, and Antragama Ewa Abbas. Thank you for the shared scientific moments during the ICT group meetings and informal discussions with great enjoyment. I also wish to express my gratitude to the helpful secretaries of the ICT Section. Laura Bruns and Jo-Ann Karna, thank you so much for your kind support and help!

Moreover, this 4-year PhD journey would not have been possible without my PhD study sponsor's financial support. Therefore, I am profoundly grateful to the Indonesia Endowment Fund for Education in collaboration with the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia (LPDP–BUDI LN). I would also like to extend my sincere thanks to Universitas Islam Indonesia for enabling and supporting me to continue my doctoral degree at the Delft University of Technology.

I am immensely grateful for the assistance given by Arthur Glenn Maail, a research manager of Open Data Lab Jakarta, to introduce my decision-making support tool and provide your best practice of opening government data in Indonesia. I also appreciate Purnama Alamsyah, a researcher at the Indonesian Institute of Science (LIPI) and a web developer of Indonesia open tender data. Purnama, I am incredibly grateful for arranging my experimental case study sessions together with the Indonesia Corruption Watch (ICW) community, policy-makers, and open data analytics.

I have had generous support from Deden Witarsyah, an associate professor and head of the Open Data Research Centre (ODRC) at Telkom University Bandung. Deden, I owe my deepest gratitude to your contribution to organising my experimental case study sessions with your colleagues, independent researchers, and students. I would also like to express my appreciation to Dedy Syamsuar to organise a focus group discussion to brainstorm open government data initiatives in the local government institution. I would like to show my most tremendous appreciation to Jon Kenedy, head of Public Information Management at the Ministry of Communication and Information Technology in Palembang, the Province of South

Sumatera. Jon Kenedy, thank you for your permission to conduct a focus group discussion and to evaluate my decision-making support tool to civil servants, decision-makers, and open government data officers.

Finally, my deep heartfelt gratitude to my family members. My mother, Umi Rusmaladewi Haroeddin, and my mother in law Ibu Mariana Minhar. Without the power of your sincere prayer, the completion of my dissertation is beyond the bounds of possibility. To my lovely wife Emigawaty, my life partner! You have always been my strength throughout the times. Thank you for your love, exponential support, and every little effort you are making for me is a memorable moment. I also dedicate this dissertation to all my beautiful children Afifa, Aqilah, and Azalia. You are the pride and joy of my life. Dear daughters, thank you for all your patience and support during my PhD journey. I would also dedicate this dissertation to my siblings Royson Firdianto, Qushoyri, and my little sister Lusida Riantimala. I owe my success to you all and thank you for your unwavering support and sacrifices. At last, I am grateful to all my brothers and sisters in law Emaliana, Siti Aminah, Deri Sugiarto, Debi Riana, Andri Syafrianto, and Hartatik.

Delft, 28 Mei 2021

List of abbreviations

ACCDB	*Access Database File*
AHP	Analytical Hierarchical Process
BbN	Bayesian-belief Networks
CCA	Creative Commons Attributions
CDF	Channel Definition Format
CSV	Comma-separated Value
DBF	Database File
DEM	Digital Elevation Model
DMS	Decision-making support
DSOD	Decision-making Support for Opening Data
DSR	Design Science Research
DSS	Decision Support Systems
DTA	Decision Tree Analysis
DTAOD	Decision Tree Analysis for Opening Data
EMV	Expected Monetary Value
ESRI	Environmental Systems Research Institute
ETL	Extract, Transform, Load
FMCDM	Fuzzy Multi-criteria Decision Making
GDPR	General Data Protection Regulation
ICT	Information and Communication Technology
JSON	JavaScript Object Notation
KKN	K-Nearest Neighbors
MANOVA	Mutivariate Analysis of Variance
MCDM	Multi-criteria Decision Making
ML	Machine Learning
NGOs	Non-government Organisations
OG	Open Government
OGD	Open Government Data
OKF	Open Knowledge Foundation
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analysis
RQ	Research Question
SC	Scientific Contribution
SLR	Systematic Literature Review
SMART	Simple Multi-Attribute Rating

Table of Contents

Summary i
Preface and acknowledgements vii
List of abbreviations x
Table of Contents xi
Table of Figures xiv
List of Tables xvi
Chapter 1 Introduction 1
1.1 The quest towards opening government data 1
1.2 Challenges to decide on opening data 2
1.3 Stakeholder tensions in the decision-making process 4
1.4 The need for decision-making support to open data 6
1.5 Problem statement and research objective 8
1.6 Research contributions 9
1.7 Structure of the dissertation 10
Chapter 2 Research approach 13
2.1 Research philosophy 13
2.2 Design science research 16
2.3 Theory building 19
2.4 Research questions 20
Chapter 3 Literature review 23
3.2 Systematic literature review approach 26
3.2.1 Data collection 28
3.2.2 Eligibility and exclusion 28
3.2.3 Deductive and inductive coding 29
3.3 Definitions and key concepts 29
3.3.1 Open government data 29
3.3.2 Advantages of opening data 33
3.3.3 Disadvantages of opening data 43
3.3.4 Elements of decision-making support 54
3.4 Taxonomy and elements of the DSOD 57
3.5 Conclusions 62
Chapter 4 Decision-making support functionalities 63
4.1 Decision-making support for opening data 63
4.2 Designing DSOD phases 65
4.3 Elements of the DSOD 67
4.4 Translation of the DSS elements to DSOD functionalities 69
4.5 Steps of the DSOD 75
Chapter 5 Methods for open data decision-making 79
5.1 Bayesian-belief Networks 79
5.1.1 Uncertainties in Bayesian-belief Networks 80
5.1.2 Structure of Bayesian-belief Networks 82
5.1.3 Conditional probability tables 84
5.1.4 Methods for the BbN’s propagation 85

5.1.5 Steps to construct the BbN 88
5.2 Fuzzy Multi-criteria Decision Making 99
5.2.1 Steps of the FMCDM 100
5.2.2 Decision alternatives 102
5.2.3 Selection criteria 103
5.2.4 Fuzzy AHP technique 105
5.2.5 Analyse dataset using FMCDM 106
5.2.6 Findings of the FMCDM steps 111
5.3 Decision Tree Analysis 113
5.3.1 Method of Decision Tree Analysis 114
5.3.2 Steps in developing the DTA 115
5.3.3 Analyse the dataset using DTA 116
5.4 A comparison of methods based on the literature 123
5.4.1 Comparison criteria 124
5.4.2 Comparison of the methods 125
5.5 Conclusions 128
Chapter 6 Development of the DSOD prototype 131
6.1 Prototyping approach 131
6.2 Prototype functionality 135
6.3 Prototype construction and demonstration 139
6.4.1 Initialisation 141
6.4.2 Analysing 147
6.4.3 Weighing 165
6.4.4 Decision-making 179
6.5 Prototype validation and testing 188
6.5.1 Alpha testing 188
6.5.2 Beta testing 188
6.5.1 Acceptance testing 189
6.6 Conclusions 189
Chapter 7 Quasi-Experiment using the DSOD 191
7.1 Evaluation methodology 192
7.1.1 Quasi-experimental approach 192
7.1.2 Structure of the quasi-experiments 193
7.1.3 Preparation of the Quasi-experiment 194
7.1.4 MANOVA analysis to evaluate the three methods 195
7.2 Quasi-experiment 1: Deciding on opening e-procurement dataset 197
7.2.1 Demographics of the participants 199
7.2.2 Experience in using open government data 202
7.2.3 Measurement of the quasi-experiment 203
7.3 Quasi-experiment 2: Deciding on opening medical records dataset 214
7.3.1 Demographics of the participants 216
7.3.2 Experience in using open government data 218
7.3.3 Measurement of the quasi-experiment 219
7.4 Comparative analysis of two quasi-experiments 228
7.5 Conclusions 232

Chapter 8 Conclusions and further research 235
8.1 Overall conclusions ... 235
8.2 Revisiting the research questions ... 238
8.3 Why decision-making processes to open data are not trivial? ... 240
8.4 Research limitations ... 242
8.5 Further research recommendations ... 243
References ... 245
Appendices ... 269
Appendix A. Preliminary Case Study ... 269
Appendix B. Experimental Case Study ... 274
Curriculum vitae ... 287
List of publications ... 289

Table of Figures

Figure 1-1 Decision-making Support to analyse government datasets 6
Figure 1-2 Structure of the dissertation 12
Figure 2-1 Design science research process 17
Figure 3-1 Literature review process and steps in this study 27
Figure 3-2 Taxonomy developed from the literature review 59
Figure 4-1 Decision-making support for opening data process 66
Figure 4-2 Translation of the DSOD functionalities 70
Figure 4-3 Translation of the functionalities and development to the DSOD steps .. 74
Figure 4-4 Steps of the DSOD 77
Figure 5-1 Basic causal structure of a BbN 83
Figure 5-2 Simple model structure illustrating nodes with two states 85
Figure 5-3 BbN before (a) and after (b) the propagation of new information 87
Figure 5-4 Step-by-step of the BbN development model 88
Figure 5-5 Identify key elements and relationships 92
Figure 5-6 Interrogating the disadvantage factors 98
Figure 5-7 The flow process of FMCDM 101
Figure 5-8 Illustration of the hierarchy of criteria and alternatives 104
Figure 5-9 Hierarchy of criteria and alternatives for the illustration 108
Figure 5-10 The pairwise comparison matrices of criteria and alternatives 110
Figure 5-11 Ranking of decision recommendations 112
Figure 5-12 Decision alternatives and possible paths 117
Figure 5-13 Decision tree analysis to estimate the costs and benefits 119
Figure 5-14 The expected monetary value and investment ranges 122
Figure 6-1 The prototyping approach and steps in this research 133
Figure 6-2 The navigation systems of the DSOD prototype 140
Figure 6-3 The main interface of the DSOD prototype 142
Figure 6-4 Sign-in process for the DSOD user 143
Figure 6-5 Message box for the user 143
Figure 6-6 Dataset selection and extraction interface 144
Figure 6-7 Dataset extraction and metadata interface 145
Figure 6-8 Dataset attribute information 147
Figure 6-9 Methods selection interface 148
Figure 6-10 Analysis steps of the Bayesian-belief Networks 149
Figure 6-11 Determine the advantage and disadvantage categories 150
Figure 6-12 Construct the causalities of the disadvantage factors 151
Figure 6-13 Construct the causalities of the benefit factors 152
Figure 6-14 Formulise the structure 153
Figure 6-15 Quantify the advantages 154
Figure 6-16 Quantify the disadvantages 155
Figure 6-17 Interrogate the disadvantages 156
Figure 6-18 Interrogate the benefit 157
Figure 6-19 Analysis steps of the FMCDM 157

Figure 6-20 Determine criteria and sub-criteria 159
Figure 6-21 Construct relationship diagram 160
Figure 6-22 Construct a hierarchical diagram 161
Figure 6-23 Define decision alternatives 162
Figure 6-24 Analysis steps of the Decision Tree Analysis 162
Figure 6-25 Determine advantages and disadvantages category 163
Figure 6-26 Construct decision structure 164
Figure 6-27 Weighing steps of the BbN 169
Figure 6-28 Assessment disadvantage and advantage factors 170
Figure 6-29 Construct matrix comparison 171
Figure 6-30 Construct pairwise comparison matrix 172
Figure 6-31 Define Fuzzy linguistic scale 174
Figure 6-32 Pairwise the advantages and disadvantages criteria 175
Figure 6-33 Pairwise advantages and disadvantages sub-criteria 176
Figure 6-34 Calculate Fuzzy weights matrix 177
Figure 6-35 Assign payoffs table 178
Figure 6-36 Construct a possible consequence 179
Figure 6-37 Decision alternatives in Bayesian-belief Networks 181
Figure 6-38 Decision alternatives 185
Figure 6-39 Rank Decision alternatives 185
Figure 6-40 Possible paths and total payoffs 187
Figure 6-41 Final decision in using DTA 187
Figure 7-1 Structure and timeline of the experiment 194
Figure 7-2 Distribution of the age of the participants 200
Figure 7-3 Type of organisation of the quasi-experiment participants 201
Figure 7-4 Current job function of the quasi-experiment participants 202
Figure 7-5 Frequency in using OGD dataset of the quasi-experiment 203
Figure 7-6 Distribution of the age of the participants 216
Figure 7-7 Type of organisation of the quasi-experiment participants 217
Figure 7-8 Current job function of the quasi-experiment participants 218
Figure 7-9 Frequency in using OGD of the quasi-experiment participants 219
Figure 7-10 Comparison of the most feasibility of the methods 229
Figure 7-11 Comparison of the methods by groups of participants 231

List of Tables

Table 2-1 Research questions (RQ) of this study 20
Table 3-1 Overview of search terms used in the systematic literature review 28
Table 3-2 Classification of Open Government Data 31
Table 3-3 Category and type of advantages in Open Government Data 34
Table 3-4 Category and type of disadvantages in Open Government Data 45
Table 3-5 Overview of DSS elements from the literature 56
Table 4-1 Elements of DSS used in this study 68
Table 5-1 Quantification of the disadvantage factors 95
Table 5-2 Row tables of Medical Records dataset 107
Table 5-3 The Fuzzy linguistic scales (adapted from: (Hsieh et al., 2004)) 109
Table 5-4 Final weights of the criteria and alternatives 111
Table 5-5 Assign payoffs and possible consequences 118
Table 5-6 Possible nodes, paths, and estimation payoffs 121
Table 5-7 Comparative the methods in opening data 125
Table 6-1 Metadata categories and their functionality 146
Table 6-2 The consequence of advantage levels to open data 166
Table 6-3 Possible occurrences and the impact of predictable benefits 167
Table 6-4 The consequence of disadvantages levels in the opening dataset 168
Table 6-5 Possible occurrences and the impact of predictable disadvantages 169
Table 6-6 The fundamental scale for pair-wise comparison (Podvezko, 2011) 174
Table 6-7 Example of weighing the advantage against the disadvantage 180
Table 6-8 The treatment plans proposed in using BbN method 182
Table 6-9 Final weights of the criteria and alternatives 184
Table 7-1 Reliability analysis of the construct of the pre-test and post-test 195
Table 7-2 Characteristics of the group participants to conduct beta testing 198
Table 7-3 Tests of normality for the transparency process 204
Table 7-4 Descriptive statistic of the transparency process 205
Table 7-5 Multivariate Tests Result for the Transparency Variable 206
Table 7-6 Tests of normality for the accuracy process 207
Table 7-7 Descriptive statistic of the accuracy process 208
Table 7-8 Multivariate Tests Result for the Accuracy Variable 209
Table 7-9 Tests of normality for the ease of use 210
Table 7-10 Descriptive statistic ease of use 210
Table 7-11 Multivariate Tests Result for the perceived ease of use Variable 211
Table 7-12 Tests of normality of usefulness 212
Table 7-13 Descriptive statistic usefulness 212
Table 7-14 Multivariate Tests Result for the usefulness Variable 213
Table 7-15 characteristics of the group participants to conduct the experiment 215
Table 7-16 Tests of normality for the transparency process 220
Table 7-17 Descriptive statistic of the transparency 220
Table 7-18 Multivariate Tests Result for the transparency Variable 221
Table 7-19 Tests of normality for the accuracy process of the DSOD prototype 222

Table 7-20 Descriptive statistic of the accuracy 222
Table 7-21 Multivariate Tests Result for the accuracy variable 223
Table 7-22 Tests of normality for the ease of use 224
Table 7-23 Descriptive statistic ease of use 224
Table 7-24 Multivariate Tests Result for the perceived ease of use variable 225
Table 7-25 Tests of normality of usefulness 226
Table 7-26 Descriptive statistic usefulness 226
Table 7-27 Multivariate Tests Result for the usefulness variable 227
Table 7-28 Overview of the comparison result between variable analysis 228
Table 7-29 Overview of the comparison result between the group of participants 230

Chapter 1
Introduction

1.1 The quest towards opening government data

Open Government Data (OGD) refers to data with legally open, machine-readable formats, non-discriminatory, and non-proprietary, which governments actively place on the OGD portals for public re-use and which can be accessed with minimal restrictions and used for free (European Commission, 2011, 2013, 2019; Geiger & Lucke, 2012; Malamud et al., 2007; Open Knowledge Foundation, 2015). The "open" definition sets out several requirements that indicate how to enable the free use, re-use, and redistribution of data (Attard, Orlandi, Scerri, & Auer, 2015; Open Knowledge Foundation, 2015). Open data refers to available data free of charge to anyone without limitations (European Commission, 2011; Kučera, Chlapek, & Nečaský, 2013; V. Wang & Shepherd, 2020; Zuiderwijk & Janssen, 2013b).

Government institutions collect and produce numerous datasets to perform their programs and tasks (Driss, Mellouli, & Trabelsi, 2019; Juana-Espinosa, 2020). These datasets' opening could improve the public's engagement, creating public trust, accountability, and transparency (McDermott, 2010). Opening public domain information through OGD initiatives can result in many advantages for society at a wide-scale (Zuiderwijk & Janssen, 2013b). Government institutions, Non-Government Organisations (NGOs), private agencies, business enablers, data enthusiasts, researchers, civil societies, and other stakeholders can benefit from opening the government's datasets (Kim, Chung, & Trimi, 2014). The advantages of disclosing data may vary, like acquiring new knowledge, creating transparency and accountability, receiving updated information about the government's ideas and achievements, generating and evaluating ideas, supporting policies and decisions, and other possible value proportions (Zuiderwijk, Janssen, & David, 2014).

The datasets' disclosure is ultimately expected to contribute to society (Ubaldi, 2013; Zuiderwijk & Janssen, 2013b). For example, parents can explore datasets about the quality of educational institutes, like secondary levels, to select a school for their children (e.g., https://www.scholenopdekaart.nl/middelbare-scholen). Researchers or academia can access various statistical datasets from an open government portal for the last five years, as they require applying for a research grant (e.g., https://dans.knaw.nl/ or https://researchdata.4tu.nl/). As the independent stakeholder reporting news, journalists might want to access the climate changes statistical datasets to explore their storylines of the news feed (e.g., https://opendata.cbs.nl/statline#/CBS/nl/).

However, deciding to open a dataset is given limited attention and encounters several challenges. Disclosing data to the public domain might face several possible disadvantages. Datasets can be used for many unknown purposes. Various types of disadvantages like personal and organisational identifiable, misuse of the data, inaccuracy of the data values, and individual discredits could trigger reluctance to open their datasets.

Furthermore, during the decision-making process to open data, stakeholders' backgrounds are often dissimilar, and their interest varies. Simultaneously, there is no decision-making support for evaluating the benefits and disadvantages of disclosing datasets. This situation results in uncertainty, which in turn might result in avoiding the opening data. Therefore, this thesis's focus is on decision-making support to help decision-makers decide on the opening datasets.

1.2 Challenges to decide on opening data

Although many datasets have been released to the public, many datasets are not fully opened (Luthfi & Janssen, 2017; Zuiderwijk & Janssen, 2015). Governments often require enormous efforts to investigate and analyse the potential disadvantages of opening datasets (Luthfi, Janssen, & Crompvoets, 2018a). The decision to either open or close the government institutions' data is not trivial (Luthfi & Janssen, 2019b).

Factors making the situation complex are the involvement of heterogeneous actors and their interests, strict regulations to process the decisions, and limited knowledge and expertise of the decision-makers (Gonzales-Zapata & Heeks, 2015;

Luthfi, Janssen, & Crompvoets, 2020). Furthermore, challenges like a lack of personal skills to analyse datasets and barriers of technology acceptance at the management level are all influencing the decision-making process to open data (Luthfi, Janssen, & Crompvoets, 2018b). Decision-makers might have insufficient knowledge and expertise to analyse the potential benefits of opening data investment and potential adverse effects like privacy violation, misuse, and ownership of the data (Luthfi, Janssen, & Crompvoets, 2019; Martin, Foulonneau, Turki, & Ihadjadene, 2013).

In addition, the decision-makers might not know which data should be opened and what decision alternatives exist besides the binary decision of "open" or "closed" the complete dataset. At the same time, there is no decision-making support for making more fine-grained decisions on whether to open or provide more suppression to the datasets before release. Furthermore, data analysts might have inadequate capabilities to retrieve, decompose, and analyse the data into an exemplary data structure. For this reason, an in-depth analysis of the datasets is needed to decide whether to open such a dataset or not.

In the open government data domain, we identified several main challenges for the stable decision-making process are identified.

1) Government agencies have limited cognitive knowledge and expertise to make decisions to open data. Agencies might have inadequate competencies to collect and analyse before releasing the datasets (Luthfi et al., 2019). There is no overview of the advantages and disadvantages of opening datasets.
2) Multi-actors decide whether to open or not to open the datasets. These actors can have different perspectives and make specific assumptions about the goals and effects of open data (Luthfi, Janssen, et al., 2018b). Different interests might stress different aspects of evaluating the impact of opening data. Hence, decision-making methods should take the diversity of interests into account.
3) Opening public and private data is a dynamic movement that may also encounter several potential disadvantages like inaccuracy, misuse, sensitivity, and inconsistency of the data (Conradie & Choenni, 2014; Martin et al., 2013). These disadvantages are often much easier to access, whereas the benefits might remain

abstract. At the same time, risk-avoiding cultures can lead the policy-makers and decision-makers to refuse to disclose more of the government's datasets (Luthfi, Janssen, et al., 2018b; Zuiderwijk & Janssen, 2017). There is a need for an overview of the advantages and disadvantages to be able to make a balance them when deciding whether to open data.

These challenges can result in the reluctance of governments to open more of their datasets. If a dataset's status is restricted and remains closed to the public, the merits of open data initiatives cannot be obtained (Luthfi et al., 2019). Furthermore, the requirements to use specific methods in the decision-making of opening data are not clear and ill-understood. Up to now, there has been no standard or procedure to analyse the advantages and possible disadvantages of opening specific datasets (Luthfi et al., 2019). Furthermore, each government institution has its unique contexts and cases on how to decide to open data.

1.3 Stakeholder tensions in the decision-making process

In the open government data domain, multiple stakeholders' backgrounds in the decision-making process are often heterogeneous. Some stakeholders set the policy, while others might want to know the progress of the current decision-making process, the time to decide, and the decisions' outcome. The policy stakeholders might want to emphasise openness, whereas decision-makers might highlight the avoidance of disadvantages as they can be held accountable when they make a mistake. However, not all stakeholders are equally involved in the decision-making process from the beginning. Some are positioned to take responsibility to direct or provide input for others to make a decision.

There are various types of stakeholders involved in the decision-making of OGD. Stakeholders in the decision-making process to open data can be primary and secondary stakeholders. The primary stakeholders include politicians, decision-makers, civil servants, and administrative officers (Gonzales-Zapata & Heeks, 2015; Zuiderwijk, Janssen, et al., 2014). While secondary stakeholders are acquired from non-governmental ecosystems, including data enthusiasts, researchers, journalists, business enablers, and civil society (Gonzales-Zapata & Heeks, 2015; M. Janssen, Charalabidis, & Zuiderwijk, 2012; Luthfi, Janssen, et al., 2018b; Zuiderwijk, Janssen, et al., 2014). The latter include the potential users of data.

The primary stakeholders refer to those data publishers with formal, official, hierarchical, and contractual relationships with the government. These primary stakeholders have a direct and essential role in deciding to open or closed the datasets. In addition, secondary stakeholders represent open data users who are influenced by the open government datasets but are less formal. These secondary stakeholders do not directly contribute to opening the data's decision-making process.

A primary challenge encountered is stakeholders having various interests and concerns in the decision-making process, like politicians, executive boards, decision-makers, civil servants, and administrative officers, all with distinct perspectives and agendas (Jetzek, Bjorn-Andersen, & Avital, 2013). These stakeholders play different roles in the decision-making process of opening data ranging from setting the objectives and ambitions to the actual opening of data. The diversity of stakeholders and their interests, various interpretations of strict regulations, limited knowledge and expertise, lack of personal skills, and barriers of technology acceptance at the management level are influencing the decision-making process of opening data (Luthfi, Janssen, et al., 2018b). As a result, the decision-making process becomes fuzzy, and the objectives of opening data are not realised. Besides, the different roles and interests of the heterogonous stakeholders in the internal government organisation might create an erratic and slow decision-making process.

For example, some decision-makers might have high authority to publish or keep the dataset closed. Some stakeholders focus on their own objectives, like the privacy officers, who should ensure that personal data will be protected when opening datasets. Furthermore, some public servants might be risk-averse, whereas others might open datasets without thinking about the possible negative consequences. In general, perceptions of the pros and cons of open data can be different among stakeholders. Furthermore, the dominating view on deciding to open data is that of being a systematic and structured process in which a careful trade-off is taken between the pros and cons of opening data, and then the best decision is made. However, reality might be more cumbersome due to the complex relationships and different stakeholders' roles and interests.

1.4 The need for decision-making support to open data

Although the potential disadvantages often dominate the decision to open data, there are no methods to decide to open datasets, including analysing the advantages and disadvantages of opening data. No comprehensive framework or a particular model exists to understand the business process resulting in disadvantages and possible adverse effects when opening data. Furthermore, it might be possible to minimise the disadvantages and other adverse effects by taking measures like reducing sensitive data elements. Hence, there is a need for decision-making support to analyse the advantages and disadvantages of opening data as visualised in Figure 1-1. Disadvantages refer to the risks of opening data and the costs involved.

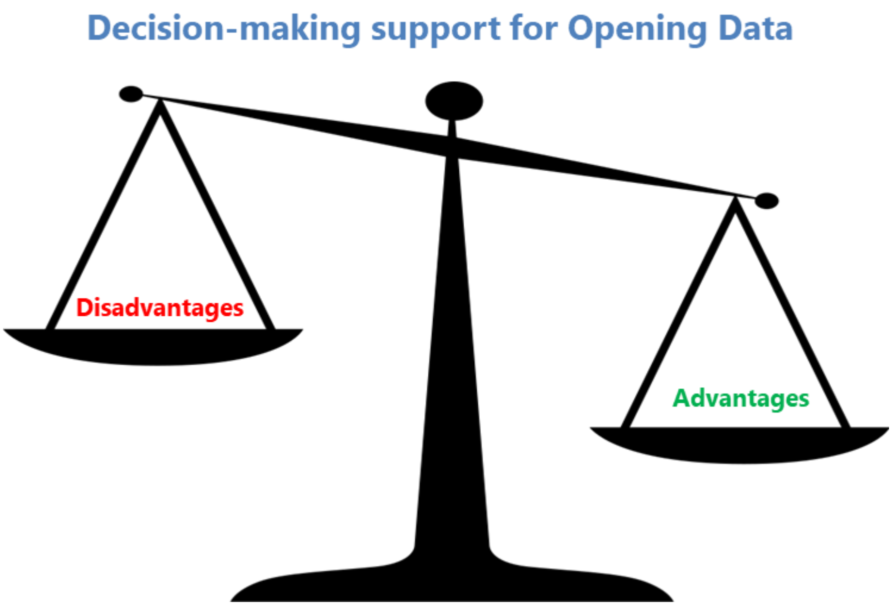


Figure 1-1 Decision-making Support to analyse government datasets

At the time we started this research, there existed no methods or proposed decision-making support for opening data in the literature. Therefore, it has not been clear what methods are the best to evaluate the potential benefits and disadvantages of the opening data. At the same time, decision-making support algorithms such as Bayesian Classifier, K-nearest Neighbors (KNN), Clustering, Fuzzy Decision Making, Analytical Hierarchy Process (AHP), and Decision Tree Analysis (DTA) are all possible methods to develop the decision-support in open data cases.

In this research, we defined the method requirements and criteria. First, since we conducted two different case studies in e-procurement and medical records case studies, the methods could be able to analyse different types of datasets. Second, we engaged various stakeholders in the experimental case studies to evaluate our proposed DSOD, such as decision-makers, politicians, civil servants, data analytics, researchers, and communities. Hence, the methods should employ the proposed DSOD by the diverse stakeholder's backgrounds and expertise. Last, because this research aims to investigate the disparate methods of transparency, accuracy, and usefulness, we picked the methods that could be compared to each other.

As a result, based on the defined requirements and its criteria, we selected three methods to analyse the benefits and disadvantages of opening datasets, namely Bayesian-belief Networks (BbN, Fuzzy Multi-criteria Decision Making (FMCDM), and Decision Tree Analysis (DTA), which are used in the analysing step. The use of the three methods in this study has different purposes and benefits. First, the BbN method allows a combination of data with domain knowledge and facilitates learning about causal relationships between variables (Heckerman, 2008). Furthermore, BbN can weigh utilities integrally against each other and consider the uncertainties in cause-effect relationships. We postulate the BbN method might be useful to organisations like government institutions with a limited number of experts to quantify the potential advantages and disadvantages of opening data.

Second, FMCDM theory in the open data domain aims to manage problems in decision-making alternatives (Ceballos, Lamata, & Pelta, 2017). The Fuzzy theory captures the experts' expertise and expresses it with the computational approach

(Ceballos et al., 2017; Rezaei, Rezaei, Nazari-Shirkouhi, & Tajabadi, 2013). The FMCDM method can likely assess the alternative selection concerning predetermined criteria for single decision-making (Kahraman, Onar, & Oztaysi, 2015). Furthermore, the FMCDM provides a very efficient solution to complex problems in the open government data field that can work with various inputs (Ceballos et al., 2017). Therefore, we postulate that this method can help decision-makers with limited experience to analyse the advantages and disadvantages of opening data.

Third, DTA is used to construct a feasible decision alternative from decision-making problems (Adina Tofan, 2015). The DTA can manage several possible costs and benefits in policy-making (Quinlan, 1990; S. Zhang, 2012). The DTA is acceptable for the quantitative and qualitative approaches when the organisations have a limited number of experts (Mittal & Khanduja, 2017). Furthermore, the DTA represents the flow of decision-making events so that decision-makers consider the uncertainty aspects and probabilities of decision outcomes (Marsh, 1993).

1.5 Problem statement and research objective

Although many datasets have been published to the public, many datasets remain closed, and datasets are often not fully disclosed. The government institutions face several challenges to release data (M. Janssen et al., 2012; Martin et al., 2013). Furthermore, stakeholders, such as politicians, executive boards, decision-makers, civil servants, data enthusiasts, and civil societies have different views on opening or disclosing datasets. Decision-makers often have no means to decide to open data, and often the potential disadvantages dominate. It is easier to keep a dataset closed rather than taking the disadvantages of the opening dataset. Therefore, decision-makers are often reluctant to open more of their datasets because of several potential disadvantages.

Furthermore, opening data to the public is a dynamic movement and might encounter several possible disadvantages. Factors like disclosing personal identity, misuse of the data, inaccuracy of the data values, and discredits of the individual or organisation could degrade the reputation of the government institutions. Furthermore, governments have limited insights and expertise in estimating possible investment costs during the decision-making process in opening data. Decision-makers and other related primary stakeholders have no supporting tools to decide whether the dataset could be opened or remain closed.

This dissertation addresses the gap that there is limited knowledge about decision-making and in analysing the potential advantages and disadvantages of disclosing specific data. Hence, we formulated the problem statement for this research as follows: *"There is limited insight and knowledge by decision-makers on how to analyse the potential advantages and disadvantages of opening data and what design of decision-making support can effectively help in deciding to open data."*

To find the solution to this research problem, we propose to develop a Decision-making Support for Opening Data (DSOD). This DSOD accommodates a systematic approach to understand decision-makers better deciding whether to open, provide limited access, introduce suppression, or remain closed the dataset to the public. Therefore, the objective of this research is "*to develop a decision-making support that provides a systematic approach to decide to open the data*."

1.6 Research contributions

Based on the previous studies, we found that there is no decision support model available for supporting the decision to open data. At the same time, opening data has specific challenges that need to be addressed by the decision-making support model. In this study, various research contributions will be made. The overall contribution is a decision-making support model, as none existed when this research was started.

First, governments and decision-makers inability to determine the potential costs, disadvantages, and advantages of opening data result in the reluctance to open more data. The first scientific contribution (SC) of this research is to provide a taxonomy of the potential advantages and disadvantages of releasing data that need to be weighed when governments want to open their data to the public (SC-1).

Second, there are no methods available to analyse the potential advantages and disadvantages of opening data. The following contribution of this research is to provide insights for decision-makers into the impact of opening datasets. Furthermore, the difference between BbN, FMCDM, and DTA, including its acceptance intervals (SC-2), will be evaluated.

1.7 Structure of the dissertation

This dissertation consists of eight main chapters. In Chapter 1, we provide the background of the study. This chapter presents the quest towards opening government data and several challenges to decide to open government data. Next, we discuss the stakeholder tensions in the decision-making process and the need for decision-making support to open data. In the last sections of this chapter, we provide the problem statement, research objective and contributions.

Chapter 2 introduces the research approach and philosophy used in this study, including the interpretivism approach. Thereafter, we present the Design Science Research methodology that we employed in this study, including the research phases. In the following sub-section, we present the research's theory-building to define the research concept's relationships. In the last section, we provide the formulation of the research questions.

In Chapter 3, we deliver the scientific literature review process to bring out the literature gaps. In this chapter, we present the data collection process, definitions and several key concepts. In the last sections of this chapter, we provide the taxonomy of the advantages and disadvantages of opening data, elements of the decision support system, and the summary of Chapter 3.

In Chapter 4, we present the systematic process towards a decision-making process to open data. This chapter uses the DSS elements derived from the literature study to define the DSOD functionalities. Next, we employ the DSR methodology to develop systematic steps of the DSOD. In the last section, we provide a summary of this chapter.

In Chapter 5, we present three methods for deciding to open data, namely Bayesian-belief Networks (BbN), Fuzzy Multi-criteria Decision Making (FMCDM), and Decision Tree Analysis (DTA). To better understand similarities and differences among the methods, we present a comparison of methods based on the literature surveys. In the last section, we describe a summary of the chapter.

In Chapter 6, we present the development of the DSOD prototype. This chapter consists of the prototyping approach, prototyping objective, prototype function selection, prototype construction, and prototype validation. In the last section, we provide a summary of this chapter.

In Chapter 7, we present the Quasi-experiment using the DSOD prototype. In this chapter, the evaluation phases consist of defining the evaluation methodology, including the quasi-experiment approach, pre-test and post-test roles, and the survey instruments' validity. To evaluate the developed DSOD prototype, we use two case studies in the domain of e-procurement and medical records.

In the last Chapter 8, we provide conclusions of the dissertation and research directions. This chapter consists of several essential elements: revisiting the research questions, why the decision-making process to open data is not trivial, research limitations, and further research recommendation at the end of the section.

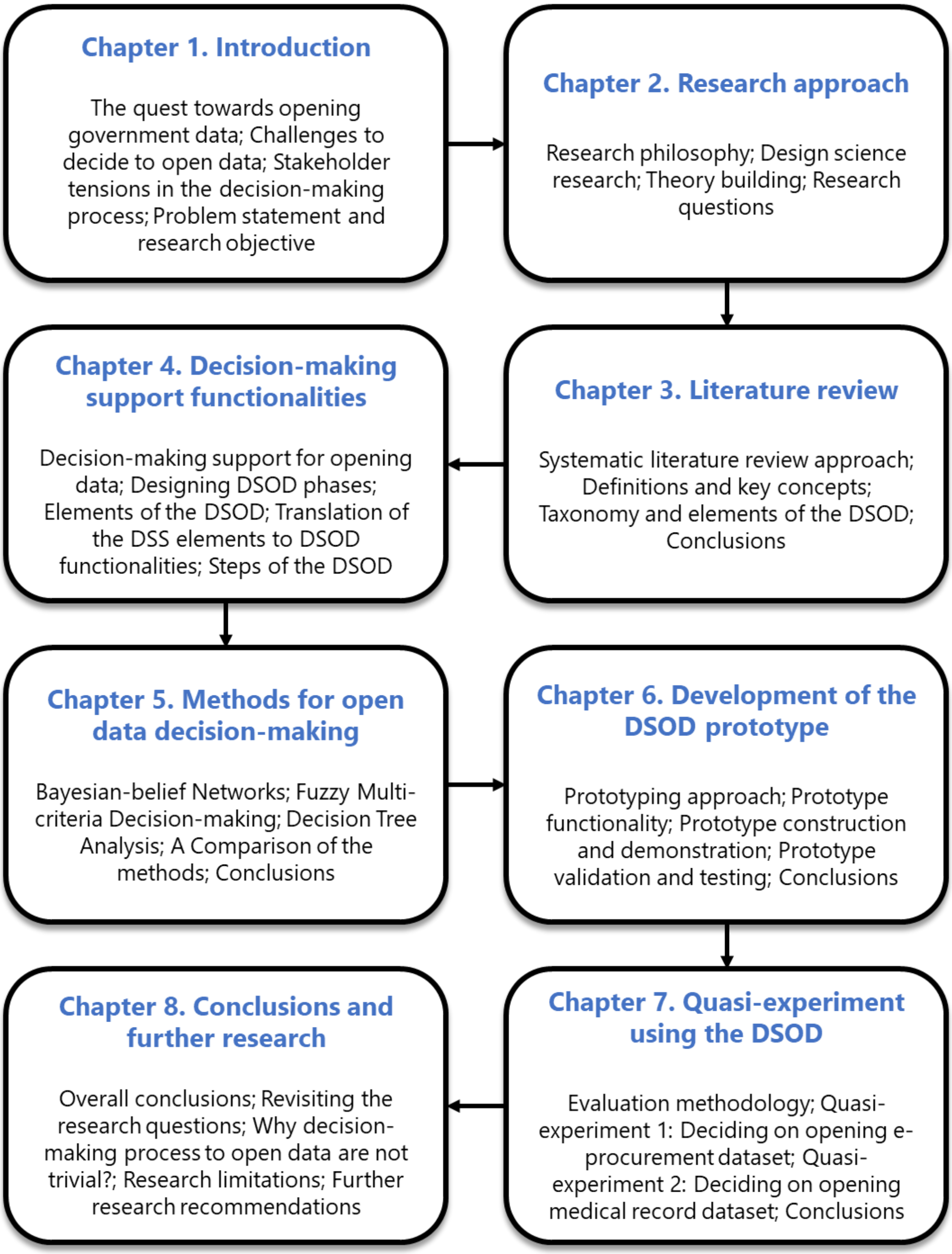


Figure 1-2 Structure of the dissertation

Chapter 2
Research approach

In this chapter, we discuss the research approach that is used to achieve the research objective. We start by discussing the research philosophy in section 2.1, followed by explaining the motivation to use design science research in section 2.2. Next, we describe how this research contributes to the theory building in section 2.3. Finally, we present the research questions in section 2.4.

2.1 Research philosophy

Research philosophy is an essential part of a research methodology (Guba & Lincoln, 1982; Holden & Lynch, 2004). The essence of research philosophy is a central belief system about how data regarding a phenomenon should be collected, analysed, and used (Guba & Lincoln, 1982). The research approach enables the researchers to decide which approach and methodology need to be adopted and also provide the motivation why a particular research method is selected (Holden & Lynch, 2004; Saunders, Lewis, & Thornhill, 2007). Therefore, before choosing the appropriate research philosophy, the researchers need to understand the philosophies in doing research (Kothari, 2004; Saunders et al., 2007).

Research philosophy is classified into three different philosophical assumptions: epistemology, ontology, and axiology (Guba & Lincoln, 1982; Holden & Lynch, 2004; Kivunja & Kuyini, 2017). Epistemology is "*the acceptable knowledge of a particular area of study*" (Saunders et al., 2007). Epistemology can be divided into two elements: resources researcher and feeling researcher (Creswell, 2014; Myers, 1997; Saunders et al., 2007). The resource researcher considers the data from the natural scientist's perspective while the feeling researcher deals with the researchers' feelings and attitudes (Creswell, 2014; Holden & Lynch, 2004; Myers, 1997; Saunders et al., 2007). The resource researcher tends to apply a positivist philosophy. Simultaneously, the feeling researcher tends to focus on interpretivism philosophy (Saunders et al., 2007). Ontology is concerned with the nature of reality that raises researchers' assumptions to understand the particular perspectives (Holden & Lynch, 2004; Kivunja & Kuyini, 2017; Saunders et al., 2007). There are two main aspects of ontological philosophy: objectivism and subjectivism (Saunders et al., 2007). Objectivism describes social entities' position, in reality, external to social actors (Saunders et al., 2007). While subjectivism portrays that social phenomena are generated from the social actors' perceptions and consequent actions (Holden & Lynch, 2004; Saunders et al., 2007). Axiology is a part of the philosophy domain dealing with three main aspects: judgment, aesthetics, and ethics (Creswell, 2014; Saunders et al., 2007). The researchers perform axiological skills to articulate values as an essential part of making judgments with regards to what research they are conducting and how they are dealing with the research (Saunders et al., 2007). The following section will describe the interpretivism philosophy further since this approach fits with this research methodology.

The use of an interpretivism approach in this research is essential to understand the decision-making process's circumstances to open data. The circumstances include the motivation and reason behind the decision-making process to open or to close the data. Furthermore, this research is also needed to understand the relationships between the advantages and disadvantages of opening data and other subjective experiences that can emerge during the observations and experimental case study.

Furthermore, several factors, such as the distinctive stakeholders' role and their interest, limited cognitive knowledge, strict regulations, and tensions at the management level, are influenced by the decision-making process to open data. (Luthfi, Janssen, et al., 2018b). As a result, the decision-making process becomes fuzzy and disclosing data is challenging to implement. Therefore, the interpretivism approach's main aim in research is to understand and interpret the meanings in human's or stakeholder's behaviour rather than generalise and predict the research variables' cause and effects (Neuman, 2014).

Interpretivism is an epistemology concerning the assessment and differences between humans as social actors (Guba & Lincoln, 1982). The interpretivism belief is that reality is heterogeneous and relative (L. A. Hudson & Ozanne, 1988). These heterogeneous realities have different meanings, making it more challenging to interpret and fix the facts (Neuman, 2014). Interpretive researchers believe that reality consists of the researcher's subjective perspectives and experiences (Black, 2006; Rowlands, 2005). In the interpretive domain, there are no correct and incorrect theories (Walsham, 1993). Therefore, the researchers adopt an inter-subjective epistemology and ontological belief when the reality is socially generated (Goldkuhl, 2012; Neuman, 2014).

An interpretive philosophy can be identified based on the assumption that knowledge of reality is obtained only through social constructions such as language, consciousness, shared meaning, documents, tools, and other research instruments (Klein & Myers, 1999). Interpretative research focuses on the complexity of human decision-making rather than predefined dependent and independent variables (Cavaye, 1996; Kaplan & Maxwell, 2005; Klein & Myers, 1999). The use of interpretation can receive several benefits (Williams, 2018). First, the interpretive philosophy is an appropriate approach to explore hidden reasons behind the complicated situation and its interrelated decision-making where quantitative evidence may be inaccurate and potentially biased. Second, the interpretive approach may be helpful for theory construction, where there is no priority theory. Third, the interpretive approach is well-suited for studying context-specific events or decision-making processes. Fourth, interpretive research-based can support finding relevant research questions and uncover decision-making issues for follow-up research.

Moreover, the interpretive paradigm is supported by observation and interpretation to collect information about events (Rowlands, 2005; Walsham, 1993). The interpretative paradigm focuses on the requirement to use analysis to comprehend individuals' subjective experiences and the relationship between the researcher and subjects (Reeves & Hedberg, 2003). The objective of interpreting the events is to give meaning to information by judging the match between the information and several abstract patterns (Aikenhead, 1997).

In this research, we acknowledge the limited cognitive capabilities of decision-makers. For example, having decision-makers with abilities in many fields like mathematics, computations, and decision-making expertise can be ideal. Nevertheless, in reality, human beings' cognitive competencies are limited (Simon, 1972). Therefore, in this research, we use the bounded rationality theory to interpret the testing results. Bounded rationality contributes to a strategy when the decision-makers rational thinking to process the decision-making to open data is limited.

2.2 Design science research

Design Science Research (DSR) is "a research paradigm in which a designer answers questions relevant to human problems via the creation of innovative artefacts, thereby contributing new knowledge to the body of scientific evidence. The designed artefacts are both useful and fundamental in understanding that problem (Hevner & Chatterjee, 2010)". The DSR aims to solve relevant classifications of problems by developing useful artefacts and constructing models, methods, and design theories (Hevner, March, Park, & Ram, 2004; Prat, Comyn-Wattiau, & Akoka, 2014). The DSR is comprehended as the method that creates, evaluates artefacts to solve problems identified in an organisation from the academic and institutional point of view (Bayazit, 2004; Hevner et al., 2004).

DSR plays an essential role in developing DSS's prototype in emerging environments (Arnott & Pervan, 2014). There are four main reasons for using DSR methodology in this study. First, DSR is focused on understanding the decision-making support requirements and decision context analysis (March & Smith, 1995). Since we develop a DSOD in the government institution context, whereby involving heterogeneous stakeholders, the DSR methodology is relevant to the governments' challenge to open their datasets. Second, the DSR methodology highlights a rigorous approach to advancing current decision-making support development knowledge (Hevner et al., 2004). As such, this is highly applicable to emerging DSOD environments, where decision-support roles involve incorporating domain knowledge for understanding the decision-making process to open or not to open the datasets. Third, the use of DSR methodology could reduce the existing gap between theory and practice (van Aken & Romme, 2012). Furthermore, this method is concerned with problem-solving and considers providing new knowledge and decision-making to decision-makers (Arnott & Pervan, 2012; Hevner et al., 2004). The DSOD is focused on developing the decision-making support steps and oriented toward the references for using the different decision-making methods to open data.

These all benefits combination and rationales in using the DSR method could guide the development of the DSOD in this study. Therefore, we use the DSR approach to develop a step-by-step decision-making process and each decision process's functionality to open data. The primary purpose of design science research is to obtain knowledge and better understand a problem domain by developing and implementing a design artefact (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007).

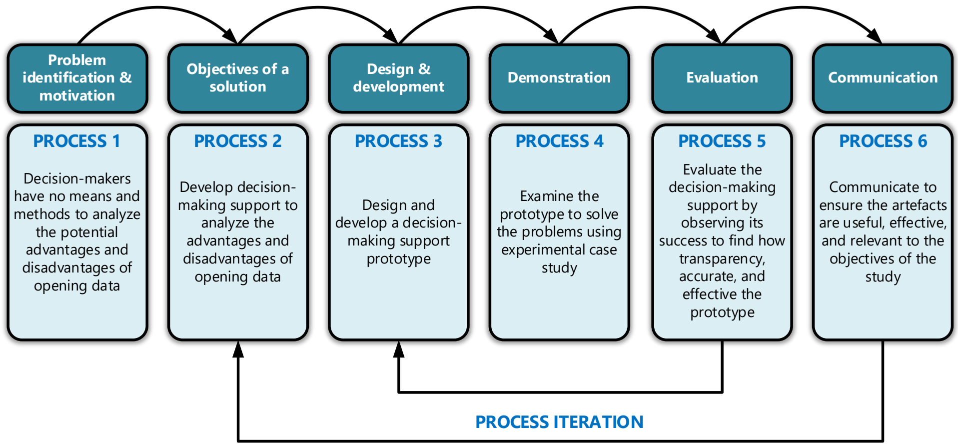


Figure 2-1 Design science research process

Figure 2-1 shows the DSR process in this research following the nominal process sequence proposed by (Peffers et al., 2007), including essential elements and their relevant problems. The detailed steps of the process can be described as follows:

1. *Problem Identification and motivation*. At this first stage, the problems are identified and defined. Government institutions face several complexities in disclosing their datasets. First, the stakeholders' involvement in the decision-making process to open data, such as politicians, executive boards, decision-makers, civil servants, data enthusiasts, and civil societies, might play different roles and views on opening and using the datasets. Second, the governments and decision-makers in particular often have no tools to support decision-making to open data. Therefore, decision-

makers are often reluctant to open more of their datasets because of the risk-averse cultures.

2. *The objective of a solution*. In the second step, the aim of the proposed DSOD is defined. This research's main objective is to develop decision-making support to weigh the advantages and disadvantages to open data. Moreover, this research's proposed DSOD is a prototype as a novel artefact in the opening data domain's decision-making support.
3. *Design and development*. In this step, systematic literature will be reviewed to provide a theoretical foundation. This foundation will serve as input to derive functionalities of the decision-support model. Furthermore, case studies will be investigated to have a deep understanding of the problem at hand in practice. The case might give more detail than literature, reveal other problems and provide a deep understanding. From the results of this literature study and the case studies, a decision-support model will be developed. In turn, a prototype will be developed to support analysing the advantages and disadvantages of opening data.
4. *Demonstration*. After designing and developing the decision-making support prototype, we demonstrate the proposed prototype's efficacy to solve the problems. To demonstrate the prototype, we use the experimental case study and some scenarios. Several resources are required for the demonstration, including decision-making support tools and participants from different organisations.
5. *Evaluation*. The evaluation stage's main purpose is to compare the predetermined solution's objectives to the actual observations. The evaluation includes comparing the three different methods to accommodate the decision-making support requirements and the diverse roles and interests of the open data's stakeholders. In this step, it is possible to reiterate the prototype development process back to step 3 to accommodate the DSOD requirements, such as user provisions and institutional regulation. At this stage, the datasets that have been selected in the previous step will go through the evaluation process. The system will interpret data that translates each data value from a table to be included in two broad categories of advantages and disadvantages.

6. *Communication*. The final stage of this process is how we communicate the issues and interests, including ensuring artefacts that have been constructed are helpful, effective, and relevant to the objectives of the study.

At the end of the design science research process, we can decide whether to iterate back step 2 (objective of a solution) to improve the artefacts' effectiveness (Peffers et al., 2007). The iteration process is also required to continue the communication and leave further improvement of the proposed DSOD prototype to a subsequent process (March & Smith, 1995; Peffers et al., 2007).

In this study, three different methods, namely Bayesian-belief Networks (BbN), Fuzzy Multi-Criteria Decision-making (FMCDM, and Decision Tree Analysis (DTA), are used in step 5 (evaluation). The methods were selected based on their variety, and differences in testing and case study performance were expected. BbN is used as it can weigh utilities between benefits and disadvantages of opening data integrally against each other and can take into account uncertainties in cause-effect relationships (Chockalingam, Pieters, Teixeira, Khakzad, & van Gelder, 2018). The FMCDM, in other methods, can provide a very efficient solution to decision-making problems in the open government data field that can work with various types of inputs such as qualitative and quantitative information from the experts. The DTA method can serve as a mechanism for dealing with the costs, benefits, and possible consequences of opening data.

2.3 Theory building

In this research, we develop a decision-making support theory for opening government data. The theory consists of several elements, including the taxonomy of the benefits and disadvantages of opening data, multiple decision-making methods, steps and rules to open data, decision-making support tools, and the causal relationships between factors influencing the advantages and disadvantages of opening data.

Isaak (1981) argued that concepts are based on generalising a number of characteristics of certain phenomena containing explanations and predictions. Judging from the definition of theory and concepts, they are related to each other. Concepts are the building blocks for the formation of a theory (Manheim & Rich, 1995). There are three reasons why the concept is considered important, namely: (1) in the empirical investigation, the concept opens the opportunity for the observation of a phenomenon; (2) to be precise; and (3) to have a theoretical impact, which is created when other concepts play an essential role in explaining the phenomena that occur (Manheim & Rich, 1995).

The DSOD is constructed using an integrated design research methodology by combining theory, system development, experimentation, and observation in a research proposal (Nunamaker, Chen, & Purdin, 1991). The development of theories, concepts, and approaches must complement each other to obtain a multidimensional and useful research model. In line with this view, several research contributions in this study consist of three main aspects. Firstly, a comprehensive overview of the potential disadvantages and benefits of opening data. Secondly, a taxonomy of the data classification is associated with the disadvantages and advantages of opening data. Thirdly, a DSOD prototype evaluates the disadvantages and advantages and the assessment of the usage by decision-makers.

2.4 Research questions

This study has formulated four research questions to develop a decision-making support theory following a design science research approach (Peffers et al., 2007). Design science research formulates the research questions to define the scope and modes of inquiry, characterise the proposed decision-support model artefact, and communicate the model to get contributions.

Table 2-1 Research questions (RQ) of this study

(RQ#1)	What are the advantages and disadvantages of opening data?
(RQ#2)	What are the elements of the decision-making support for opening data?
(RQ#3)	What are the functionalities of a prototype?
(RQ#4)	What are the differences between BbN, FMCDM, and DTA to support decision-making about opening the dataset?

The first research question (RQ#1) seeks to answer the advantages and disadvantages of opening data. These advantages and disadvantages of opening data will be studied by conducting a comprehensive literature review to overview the current issues. The answer to this question should contribute to developing a taxonomy of the advantages and disadvantages of opening data, which can be used as the basis for a decision-making support prototype.

The investigation results of the first research question will contribute to the next step to identify what elements can be used to develop decision-making support (RQ#2). The elements of the decision support model will be identified by using the literature review. Also, we perform a case study to understand the decision support requirements from the decision-makers perspective. The literature study and performance of the case studies will result in a list of decision-making support elements and detailed steps that need to be taken to make decisions about opening data.

The decision support model's prototype is developed in the third research question (RQ#3). The taxonomy of advantages and disadvantages in RQ#1, the decision support elements in RQ#2, will be used as the basis for defining the proposed decision support model's functionalities. We will design the testing approach to indicate the developed prototype's performance from the proposed decision support model's functionalities.

The fourth research question (RQ#4) seeks to answer the BbN, FMCDM, and DTA differences to support decision-making open data. To answer RQ#4, we experimented by using three different groups of stakeholders. The stakeholders are representing governments, academia/universities, and communities/professionals. There are three main factors to measure the differences between the three selected methods. The factors include how transparent the process is, how accurate the expected results are, and how useful the proposed DSOD is for the open government data stakeholders.

Chapter 3
Literature review

This chapter presents a literature review as the basis for this research. The objective is to answer the first (RQ#1) and the second (RQ#2) research question. The instrument used in this study is a systematic literature review (SLR), which consists of a lithographic overview. There are two main objectives to do a systematic literature review in this chapter. The first objective is to define this research's main concepts, including several terminologies like decision-making support and the advantages and disadvantages of opening data. This literature study's result is the answer to RQ#1, *what are the advantages and disadvantages of opening data*? Subsequently, we use a literature study to define the decision-making support elements, including its requirements and the detailed step of making open data decisions. This literature survey is the answer to the second research question (RQ#2), *what are the elements of decision-making support opening data?* Parts of this chapter have been published in (Luthfi & Janssen, 2017, 2019a; Luthfi, Janssen, et al., 2018a, 2018b).

3.1 Literature review approach

Literature reviews play an essential role in all research disciplines and research projects (Paré & Kitsiou, 2015; Snyder, 2019). A literature review can be delineated as a systematic approach to gathering and synthesising prior research (Baumeiester & Leary, 1997). A well-conducted literature review process as a research method by integrating the study's perspectives and findings can address research questions (Snyder, 2019; Webster & Watson, 2002). The literature review's objective is to help readers understand the entire body of available research on a specific topic (De Los Reyes & Kazdin, 2008; Rhoades, 2011). The literature review approach can also help provide the study domain's strengths, weaknesses, and potential gaps (De Los Reyes & Kazdin, 2008).

In this research, the literature review approach is used to attain several purposes. First, it provides the study findings and results strictly related to the study being reported (Fraenkel & Wallen, 2011). One of the main objectives of carrying out a literature review in this study is to systematically explore the research's existing position in the decision-making support to open data by investigating the prior studies. Second, it connects a study of ongoing discussions in the literature about a topic and determines gaps and possible blank space from the previous research (C. Marshall & Rossman, 2010). This research also considers providing the possible similarity and diversity of the key terminologies such as open government data, decision-making support, advantages, and disadvantages of opening data. Therefore, we give attention to generate the gaps among these critical terminologies. Third, it provides a conceptual model or framework for establishing the research's essential scientific contribution (Denney, 2013). This research uses a systematic literature review approach to generate a taxonomy of the advantages and disadvantages of opening data to answer the first research question (RQ#1). Also, the review was used to derive the elements of the decision-making process for opening data to answer the second research question (RQ#2).

There are several types of literature review approaches and methods for gathering and synthesising existing literature, namely:

1. *Narrative reviews*

   The narrative review is a conservative method of reviewing the existing literature source and uses a qualitative approach to interpret the previous knowledge and study domain (Sylvester, Tate, & Johnstone, 2010). This approach often synthesises and interprets the literature to demonstrate the knowledge domain's specific perspective (Baumeiester & Leary, 1997). Therefore, narrative reviews tend to use an unsystematic approach (Paré & Kitsiou, 2015). As a result, the primary articles' selection of information is subjective (Paré & Kitsiou, 2015; Snyder, 2019). Thus, this approach can lead to a biased interpretation and lacks explicit relevant criteria for making references (B. N. Green, Johnson, & Adams, 2006).

2. *Descriptive reviews*

   A descriptive review's main objective is to determine the level of body knowledge in a specific domain of research topic exposes with regards to existing propositions, theories, methodologies, and research findings (King & He, 2006; Paré & Kitsiou, 2015). Contrary to the narrative reviews approach, descriptive analyses adhere to a systematic and transparent way of searching, filtering, classifying, and analysing studies (Petersen, Vakkalanka, & Kuzniarz, 2015). A descriptive review approach also extracts some essential elements related to the article characteristics like the publication year, research methods, data collection techniques, and research outcomes' strengths and weaknesses (Sylvester et al., 2010).

3. *Systematic reviews*

   A systematic review aims to identify and critically appraise relevant research topics and collect and analyse data from prior research (Snyder, 2019). A systematic review aims to identify empirical evidence suitable for relevant articles' criteria to answer the specific research questions (Paré & Kitsiou, 2015). Therefore, the potential bias can be minimised by performing the systematic approach while reviewing articles (Moher, Liberati, Tetzlaff, & Altman, 2009). Furthermore, a systematic review process can use compassionate and structured strategies to identify relevant studies, both non-published and published articles (Moher et al., 2009).

In this research, we used SLR to synthesise empirical evidence to answer the first and second research questions. Taking a systematic review approach in this research has several benefits (Moher et al., 2009; Paré & Kitsiou, 2015). First, an explicit method that enables existing aggregate research in decision-making support to open government data. Second, the systematic review is used to assess the relationships between the advantages and disadvantages of opening data. Third, systematic reviews can identify and explain the consistencies between study results and define the different research outcomes.

3.2 Systematic literature review approach

Systematic review approaches have obtained substantial attention in most study domains over the years (Berrang-Ford, Pearce, & Ford, 2015). A systematic review integrates a structured and transparent process during the data collection while considering the rigorous analysis (Attard et al., 2015; Gough, Oilver, & Thomas, 2012). This section describes the literature review method, which follows the three main sequential steps of the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) Protocol, namely data collection, eligibility and exclusion, and deductive and inductive coding (Biesbroek et al., 2018; Moher et al., 2009), as can be seen in Figure 3-1. The objective of using Prisma Protocol in this study is to ensure the quality of the literature review process by offering an efficient process (Moher et al., 2009).

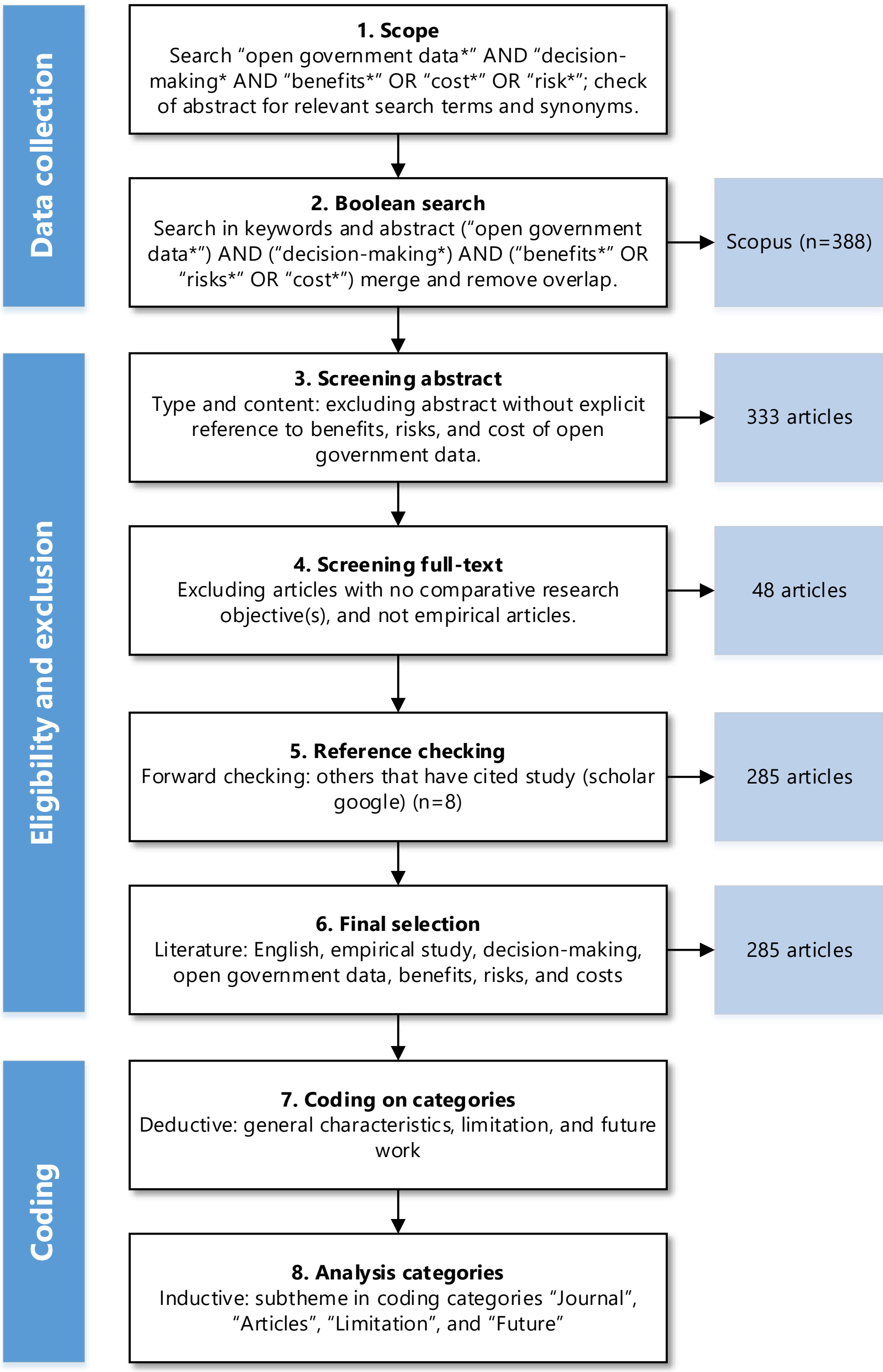


Figure 3-1 Literature review process and steps in this study

3.2.1 Data collection

To capture the relevant articles that meet within this research scope, we conducted an initial scoping of literature reviews to define and identify proper essential search terms (Biesbroek et al., 2018). This study used the Elsevier Scopus database as a well-established abstract and citation database (Harzing & Alangkas, 2016) with enriched data. Later, we linked scholarly content to cover both topical and non-topical articles. Articles were selected for the period from January 2005 to August 2020. Based on the Boolean search, we found (n=388) eligible articles represented the key construct of "open government data" AND "decision-making*" AND "benefits" OR "risks*" OR "costs*". The overview of search terms in the systematic literature review can be seen in Table 3-1.

Table 3-1 Overview of search terms used in the systematic literature review

Key Construct	Search strings
OGD	Open data, open government data, public data, big open data
Decision	Decision-making*, decision-making support*, policy-making*
Benefit	Benefit*, merits*, value*, profit*, advantage*
Risk	Risk*, endanger*, threat*, jeopardy*, disadvantage*
Cost	Cost*, revenue*, amount*, expense*, fee*, charge*

3.2.2 Eligibility and exclusion

In this step, the inclusion was limited to English-language scientific articles and empirical articles using Scopus based online scientific database. The Scopus database was searched on the "abstract" to ensure that the content was aligned with our literature objectives rather than to find unnecessary words or contents (Biesbroek et al., 2018). Furthermore, we filtered our previous articles by screening the abstract without explicit reference to benefits, risks, and costs from 388 eligible articles to (n=333). In the next step, we filtered the full text by excluding articles with no comparative research objectives and not empirical papers. Applying these criteria stepwise, we then found (n=48) eligible articles.

Furthermore, we used forward and backward reference checking or chain checking (Biesbroek et al., 2018; Gough et al., 2012). Forward checking aims to identify and examine articles that refer to our sample articles (Baumeiester & Leary, 1997). The objective of backward checking is to identify articles that are included in the reference list of the 48 articles in the screening text step (Berrang-Ford et al., 2015; Biesbroek et al., 2018). We used the Google Scholar platform to capture relevant articles based on their title, bringing the final selection to 285 articles.

3.2.3 Deductive and inductive coding

We developed a coding mechanism and data extraction table to synthesise the literature review results in the third step. The main categories consist of descriptive information ("year", "journal", "scale"), study design ("sampling frame", "study design", "data source"), limitation of the articles, and future research works (Baumeiester & Leary, 1997; Biesbroek et al., 2018; Gough et al., 2012). The deductive information from the selected articles from the second steps (eligibility and exclusion) was extracted using a data extraction table by synthesising the categories, especially the "limitation" and "future research" (Biesbroek et al., 2018; B. N. Green et al., 2006). Next, the articles' inductive information was classified based on the conceptual, empirical, and methodological limitations (Baumeiester & Leary, 1997; Berrang-Ford et al., 2015).

3.3 Definitions and key concepts

The next step is to define the key concepts of this study. The following sections discuss three main topics: Open Government Data, advantages and disadvantages of open government data.

3.3.1 Open government data

Throughout this thesis, several key concepts and terms are used. At the outset, these concepts and terms are introduced and defined as follows:

First, Open Government. Open Government (OG) is an evolving strategy for changing how governments communicate with their citizens by using Information and Communication Technologies (ICT) in more innovative and solid ways (Malamud et al., 2007; Meijer & Thaens, 2009; Witarsyah Jacob et al., 2017). Open government strategy enables government institutions to seek help and support from their citizens whenever needed to solve specific problems (Bertot, Jaeger, & Grimmes, 2010; McDermott, 2010). Therefore, the open government movement is trying to improve a more effective organisation and a more robust democracy sense (Alonso, 2011; Geiger & Lucke, 2012).

Second, Open Data. Open Data is a specific type of data regarded as 'open'. It is available free of charge for everyone to access, use, re-use, and redistribute without any restrictions (Davies, 2010; Gurstein, 2011; Zuiderwijk & Janssen, 2013b). The data should be provided altogether, preferably downloadable via the Internet, and any additional information needs to adhere to the open data license' regulation (van Loenen, Vancauwenberghe, Crompvoets, & Dalla Corte, 2018). The publishers' data should be made available in both human-readable and machine-readable formats without personal or sensitive information (Malamud et al., 2007). If the government has generated data that is available to the public domain in accordance with Open Data principles, it is referred to as Open Government Data (Kučera et al., 2013; Zuiderwijk & Janssen, 2013b).

Third, according to the Open Knowledge Foundation (OKF), Open Government Data is information that is collected and produced by the government, which is 'open' in the sense that it can be freely accessed, used, re-used, and distributed by anyone (Gigler, Rahemtulla, & Custer, 2011; K. Janssen, 2011; M. Janssen et al., 2012; Open Knowledge Foundation, 2015). Opening public domain information through open government data (OGD) initiatives can result in many advantages for society (Zuiderwijk & Janssen, 2013b). Government institutions, non-government organisations (NGOs), private agencies, business enablers, data enthusiasts, researchers, civil societies, and other stakeholders can benefit from opening the government's datasets (Kim et al., 2014). The advantages of disclosing data may vary, like acquiring new knowledge, creating transparency and accountability, receiving updated information about government's ideas and their achievements, generating and evaluating ideas, supporting policies and decisions, and other possible value proportions (Scholl & Luna-Reyes, 2011; Zuiderwijk, Janssen, et al., 2014).

This study defines seven groups of classification and perspectives regarding Open Government Data, as can be seen in Table 3-2.

Table 3-2 Classification of Open Government Data

	OGD classification (∑n=48)	Definition	Source
1	General overview *(n=8)*	This perspective advocates the benefits of opening government data for organisations like government institutions, business enablers, entrepreneurs, academia, and professionals.	(Charalabidis, Loukis, & Alexopoulos 2014; Hielkema & Hongisto, 2013; Ivanov, Varga, & Bach, 2014; Kuk & Davies, 2011; Serra, 2014; van Loenen et al., 2018; Zuiderwijk, Choenni, Janssen, & Meijer, 2014; Zuiderwijk & Janssen, 2013a)
2	Political and policy-making *(n=8)*	This perspective deals with problems implementing open government policy. The issues are derived from political issues, different interests, role gaps, security, and data protection.	(Bates, 2014; Conradie & Choenni, 2014; Dulong de Rosnay & Janssen, 2014; Kassen, 2013; Kulk & van Loenen, 2012; Linders, 2013; Yannoukakou & Araka, 2014; Zuiderwijk, Gascó, Parycek, & Janssen, 2014)
3	Organisational and institutional (*n=6)*	Organisational and institutional perspective demonstrates open government data to enable and adapt the public data as a necessary process and not as discreet and irregular.	(Andreoli-Versbacha & Mueller-Langera, 2014; Dulong de Rosnay & Janssen, 2014; Estermann, 2014; M. Janssen et al., 2012; McDonald & Léveillé, 2014; Zuiderwijk, Janssen, et al., 2012)
4	Social and cultures (*n=7*)	The social and cultural perspective brings open	(Alexopoulos , Zuiderwijk, Charapabidis,

	OGD classification (∑n=48)	Definition	Source
		government data to focus more on the merits and challenges of using government data.	Loukis, & Janssen, 2014; Bichard & Knight, 2012; Garbett, Linehan, Kirman, Wardman, & Lawson, 2011; M. Janssen et al., 2012; Jetzek et al., 2013; Josefin Lassinantti, Bergvall-Kåreborn, & Ståhlbröst, 2014; Zuiderwijk & Janssen, 2014)
5	Economic and innovation (*n*=6)	This perspective demonstrates that innovation and new ideas from open government data initiatives can stimulate economic growth. The prospectus of opening data may be able to analyse the costs-benefits and revenue stream of an organisation.	(Craveiro, Porto De Albuquerque, & Tavares de Santana, 2013; M. Janssen et al., 2012; Jetzek et al., 2013; Kassen, 2013; Josefin Lassinantti et al., 2014; Lindman, 2014)
6	Technical and human's cognitive *(n=7)*	The technical and human cognitive perspective demonstrates how people deal with the practical aspects of analysing the government's data. The technical issues may consist of data analysis techniques, data visualisation, and data treatment.	(Alexopoulos et al., 2014); Behkamal, Kahani, Bagheri, and Jeremic (2013); (Borglund & Engvall, 2014; Chan, 2013; Fleisher, 2008; Yoose & Perkins, 2019; Zuiderwijk & Janssen, 2014; Zuiderwijk, Jeffery, & Janssen, 2012a, 2012b, 2013)
7	Legislation *(n=6)*	This perspective shows the legal issues from open government data consequences (e.g., the potential conflict of existing regulations; pros and cons of policies).	(Catherine, 2012; Dulong de Rosnay & Janssen, 2014; M. Janssen et al., 2012; Kassen, 2013; Krotoski, 2012; Tsiavos, Karounos, & Stefaneas, 2013)

3.3.2 Advantages of opening data

Over the last decade, there has been a significant movement of open government data initiatives to create transparency, accountability and stimulate citizen engagement (Rui Pedro Lourenço, 2015; Luthfi & Janssen, 2017; Zuiderwijk & Janssen, 2013b). The merits of opening data like to enhance trust, improve credibility and reputation are the main drivers of government institutions to open more their data (Ali-Eldin, Zuiderwijk, & Janssen, 2017; Putri Nugroho, Zuiderwijk, Janssen, & de Jong, 2015; Zuiderwijk & Janssen, 2013b). In addition, the public expects governments to disclose more of their datasets for various kinds of purposes. Providing accessible and available datasets is a strategy to help government institutions become more credible and subsequently enhance interaction with stakeholders (Pereira, Macadar, Luciano, & Testa, 2017). For these reasons, the disclosure of datasets to the public is ultimately expected to improve the decision-making process by government institutions, business enablers, and individuals (Ubaldi, 2013; Zuiderwijk & Janssen, 2013b).

Table 3-3 Category and type of advantages in Open Government Data

	Advantages category (∑n=150)	Type of advantage	Brief description	Source
1	**Political and legislation** *(n=50)*	1.1 Improved transparency *(n=7)*	Sharing of the datasets will increase the transparency of the government and individual performance. Society is being able to access the proper information through precise datasets. This situation can improve the decision-making process and could save the investment of money both by the government and society	(Cucciniello, Nasi, & Valotti, 2012; Gigler et al., 2011; M. Janssen et al., 2012; Kucera & Chlapek, 2014; Rui Pedro Lourenço, 2013; Rui Pedro Lourenço, 2015; Saxena & Muhammad, 2017)
		1.2 Enhanced accountability *(n=6)*	The accountability of the impact of data disclosure may also influence the public's flexibility to process reliable information. Institutions or public service providers will be easy to choose which datasets they need.	(Gigler et al., 2011; M. Janssen et al., 2012; Kučera et al., 2013; Rui Pedro Lourenço, 2015; Mayernik, 2017; Saxena & Muhammad, 2017)
		1.3 Political awareness *(n=5)*	Open Government Data is considered to have situated the use of ICT-based and new technologies to stimulate data sharing in the context of political accountability and political awareness. Hence, the OGD initiatives obscure the difference between the technology-based opening data and the open government's politics.	(Attard et al., 2015; M. Janssen et al., 2012; Josefin Lassinantti et al., 2014; Puron-Cid, Gil-Garcia, & Luna-Reyes, 2012; Ubaldi, 2013)

	Advantages category (∑n=150)	Type of advantage	Brief description	Source
		1.4 Improved policy-making *(n=7)*	Opening data creates a high level of confidence for the policy-makers to share their data. Therefore, users can verify and validate the data to generate their new policy-making or sharpen policy-making alternatives.	(Attard et al., 2015; M. Janssen et al., 2012; Jetzek, Avital, & Bjorn-Andersen, 2014; Linders, 2013; Safarov, Grimmelikhuijsen, & Meijer, 2017; Veljković, Bogdanović-Dinić, & Stoimenov, 2014; Zuiderwijk, 2017)
		1.5 Increased reputation *(n=5)*	External organisations can collaborate with government agencies since the governments provide free access to the government's data. Thus, one of the main impacts of the disclosure data is that reputation can guarantee companies' long-term contracts with the governments.	(Arcidiacono & Reale, 2016; Carter, 2012; Klabi, Mellouli, & Rekik, 2018; van Loenen et al., 2018; Zuiderwijk & Janssen, 2013b)
		1.6 Improved government services *(n=6)*	Data availability of government services improves their accessibility and helps citizens and organisations utilise them better. If the users can provide feedback about the published datasets, they might notify the curators of these datasets about possible errors in data.	(Archer, Dekkers, Geodertier, Hazard, & Loutas, 2013; Barnickel et al., 2012; M. Janssen et al., 2012; Kucera & Chlapek, 2014; Schwegmann, 2012; Zuiderwijk, Janssen, et al., 2014)

	Advantages category (∑n=150)	Type of advantage	Brief description	Source
		1.7 Data sharing agreement *(n=4)*	Data sharing agreement is the machine-readable protocol for regulating data sharing among different organisation levels. The role of a data-sharing agreement in opening government data is to manage the government's information supply chains to ensure that their data is adequately protected.	(Caimi, Gambardella, Manea, Petrocchi, & Stella, 2015; Costantino, Martinelli, Matteucci , & Petrocchi, 2017; Fran. Ruiz et al., 2016; Swarup, Seligman, & Rosenthal, 2006)
		1.8 Evidence-based policy *(n=10)*	Evidence-based policy-making is a strategy that uses evidence in the core position of the policy-making process to improve decision-making more efficient and effective. Evidence-based policy-making in the opening data decision can help decision-makers make well-informed decisions by placing the best available evidence from the research repositories.	(Greenhalgh & Russel, 2009; Head, 2010; Luthfi & Janssen, 2019b; Monroe, 2011; Straßheim & Kettunen, 2014; Strydom, Nortje, Funke, & Steyn, 2010; Sundell, Tengvald, Soydan, & Anttila, 2009; Sutcliffe & Court, 2005; Urahn, Caudell-Feagan, & Stasch, 2014; Volmink, 2017)
2	**Technology** *(n=39)*	2.1 Linked open data *(n=8)*	Linked open data means that the data is well-structured and released according to the principles of linked data. The principles include that data should be interconnected, accessible, and shareable through the semantic web. The data providers can use this concept to	(Archer et al., 2013; Dunsire, 2013; Geiger & Lucke, 2012; M. Janssen & Van den Hoven, 2015; Khusro, Jabeen, Rahman Mashwani, & Alam, 2014; Yoose & Perkins, 2019;

Advantages category (∑n=150)	Type of advantage	Brief description	Source
		connect the government's datasets from different sources, making it useful for more stakeholders.	Zuiderwijk, Jeffery, et al., 2012a, 2012b)
	2.2 Data optimisation *(n=4)*	Data optimisation aims to manage data in a way that improves the quality of the released datasets. The optimisation can help the decision-makers to accelerate their decision-making process. Besides the decrease in processing speed when analysing the datasets, it is possible to reduce the overall cost instead of using traditional data analysis.	(Buneman, Davidson, & Hillebrand, 1996; Emrouznejad, 2016; M. Janssen et al., 2012; Roy, Swarup Rautaray, & Pandey, 2018)
	2.3 Data exploration *(n=4)*	Data exploration uses visual-based exploration to understand better what is in a dataset and the released datasets' characteristics. Data exploration might help the governments to complete, correct, and relate among the datasets. Hence, the data users can define metadata's basic concepts such as structure, statistics, and relationships for further analysis.	(Choe, Lee, Zhu, Henry, & Baur, 2017; Deligiannidis, Kochut, & Sheth 2007; Idreos, Papaemmanouil, & Chaudhuri, 2015; Keim, 2001)

Advantages category (∑n=150)	Type of advantage	Brief description	Source
	2.4 Data discovery *(n=5)*	Data discovery aims to describe collecting data from various sources. The governments provide visual navigation data to ease the users in advanced analytics of the published datasets.	(Gregory, 2020; M. Janssen et al., 2012; Korba et al., 2008; Weikum, 2013; Wu, Psomopoulos, Jodha Khalsa, & de Waard, 2019)
	2.5 Data validation *(n=4)*	Data validation ensures that data have undergone data cleansing, correct, useful, and high quality. Data validation strategies can avoid out-of-range data entry errors and the potential of streamlining data elicitation. The data publishers provide a well-defined, accurate, and consistent form for any kind of input from open data users. The open data users can check that the data is correct and get insight into the possibility of data conflicts.	(Davies, 2010; Gao, Xie, & Tao, 2016; Gibson, Ramwell, & Day, 2016; Horn, Miksch, Egghart, Popow, & Paky, 1998; Kupzyk & Cohen, 2015)
	2.6 Data combination *(n=5)*	Releasing datasets to the public can provide supplemental and more accurate information on the published datasets. The open data users can use statistical methods for combining multiple data sources to generate a record linkage among the datasets.	(M. Janssen et al., 2012; Komarova, Nekipelov, & Yakovlev, 2018; Nkurunziza, 2019; Wilson, Graves, Hamada, & Reese, 2006; Zuiderwijk, Jeffery, et al., 2012a, 2012b)

	Advantages category (∑n=150)	Type of advantage	Brief description	Source
		2.7 Data machine-readable *(n=9)*	One of the main characteristics of the open government data initiative is that the released data should be in a machine-readable format. The machine-readable format allows open data users to process the dataset using multiple file formats such as CSV, JSON, XML, and GTFS.	(Alonso, 2011; Ariss, 2017; Davies, 2010; Goëta & Davies, 2019; M. Janssen et al., 2012; Jetzek et al., 2013; Julia Zhu & Freund; Malamud et al., 2007; Ubaldi, 2013)
3	**Social** *(n=43)*	3.1 Improve citizen engagement *(n=6)*	Community engagement means a dynamic relational process that facilitates interaction, involvement, and communication exchange between an institution and a social community for better outcomes	(Campos & Evans, 2013; Canares, Marcial, & Narca, 2016; Gurin, 2014; Johnston, 2007; Ubaldi, 2013; H.-J. Wang & Lo, 2016; Zuiderwijk et al., 2013)
		3.2 Data reusability *(n=5)*	The data collected in the government data portal is useful for the public to expose variability and enable experimentation. The public can re-use the data to generate new ideas or knowledge based on the experimental data	(Barnickel et al., 2012; Chan, 2013; K. Janssen, 2011; Mayinka et al., 2013; Vetrò et al., 2016)
		3.3 Better decision-making *(n=5)*	Open government data initiatives provide new insights that data can be distributed, communicated, and share with the broader public domain. These data can serve as input for society to	(M. Janssen et al., 2012; G. Lee & Kwak, 2012; Linders, 2012; Sunderberg, 2016; Zuiderwijk & Janssen, 2013b)

	Advantages category (∑n=150)	Type of advantage	Brief description	Source
			make informed and better decision-making.	
		3.4 Combating corruption *(n=5)*	Open government data can help combat corruption through accountability and generate novel applications that promote public service transparency. The availability of the data in the government portal would help reduce potential corruption by increasing transparency	(Attard et al., 2015; Bertot et al., 2010; M. Janssen et al., 2012; Máchová, 2007; Rajshree & Srivastava, 2012)
		3.5 Improve public trust *(n=6)*	The public sector can use open data to inform citizens about their actions better. By publication of open data, a public-sector body can present itself as an open and transparent institution.	(M. Janssen et al., 2012); (Kucera & Chlapek, 2014); (G. Lee & Kwak, 2012; Piotrowski & Ryzin, 2007; Schwegmann, 2012; Ubaldi, 2013)
		3.6 Increase public satisfaction *(n=7)*	Data disclosure is often in line with the many promises like improving transparency, accountability, and enhancing citizen engagement. Open government data can potentially increase citizen satisfaction by providing accuracy, fairness, understandably, and ease of use to the released datasets.	(Albano & Reinhard, 2014; Cheol Kim & Yong Gim, 2015; Davies, 2010; Helbig, Cresswell, Burke, & Luna-Reyes, 2012; Kucera & Chlapek, 2014; Napolitano, 2019; Zuiderwijk & Janssen, 2013b)
		3.7 Data availability *(n=5)*	The availability of the data refers to the process of ensuring that data is available to end-users without restriction.	(M. Janssen et al., 2012; Meijer & Thaens, 2009; Pasquetto, Randles, &

	Advantages category (∑n=150)	Type of advantage	Brief description	Source
			Providing high data availability can accelerate stored data is accessible to anyone and valid in the real-time process.	Borgman, 2017; Reggi, 2011; Tanaka, 2016)
		3.8 Encourages public literacy *(n=4)*	Open government data initiatives advocate support to the citizen by enabling public information. The public can read, write, and communicate the data in specific contexts by reusing the available datasets. Publishing the full version of open datasets to society will significantly affect the willingness to create, re-use, and analyse the datasets.	(Abella, Ortiz-de-Urbina-Criado, & De-Pablos-Heredero, 2019; Elena, López , Paciello , & Pane, 2016; Kučera et al., 2013; Napolitano, 2019)
4	**Economic** *(n=26)*	4.1 Knowledge-economy growth *(n=6)*	Business process models for opening data have emerged in response to the knowledge-economic opportunities presented by the increasing availability of open data by governments. Open government data enables greater transparency and accountability, delivery of new business ideas, and stimulation of open innovations in government organisations and business enablers.	(Gonzales-Zapata & Heeks, 2015; Jetzek et al., 2014; Jetzek et al., 2013; Magalhaes, Roseira , & Strover, 2013; Zeleti , Ojo, & Curry, 2016; Zuiderwijk & Janssen, 2017)
		4.2 Innovation *(n=6)*	The sophistication of application innovation in open data systems makes it possible to introduce potential	(M. Janssen et al., 2012); (Kucera & Chlapek, 2014);

	Advantages category (∑n=150)	Type of advantage	Brief description	Source
			investment. The openness of data can improve understanding of how to process data and project it properly	(Chan, 2013; Jetzek et al., 2014; Schwegmann, 2012; Yang & Kankanhalli, 2013)
		4.3 Increased efficiency (*n=5)*	Open government data is publicly available, non-confidential and non-privacy restricted and freely available, and redistributed without additional cost.	(Alzamil & Vasarhelyi, 2019; Gonzales-Zapata & Heeks, 2015; M. Janssen & Van den Hoven, 2015; Liang, 2012; Susha, Grönlund, & Janssen, 2015)
		4.4 New business opportunity *(n=5)*	Opening data to the public can encourage both governments and society to access and re-use the data. Some potential impacts can be gained, like creating a new product, improving government services, and developing a new business model.	(M. Janssen et al., 2012; Jetzek et al., 2014; Jetzek et al., 2013; Kucera & Chlapek, 2014; Parycek, Hochtl, & Ginner, 2014)
		4.5 New jobs creation *(n=4)*	Open government data has been looked at as the primary driver of public innovation and co-creation. The availability of valuable datasets is believed can generate many more benefits. The merits include economic growth, new business and products, public services, revenue streams, and new job creation.	(Attard et al., 2015; Gonzales-Zapata & Heeks, 2015; Kucera & Chlapek, 2014; Toots, McBride, Kalvet, & Krimmer, 2017)

3.3.3 Disadvantages of opening data

Although initiatives to open data can create many benefits, they might also create disadvantages (Luthfi & Janssen, 2017; Zuiderwijk & Janssen, 2015). Disadvantages and risks are closely related. *Risks* refer to the chance these disadvantages come true and their impact (Kucera & Chlapek, 2014; Luthfi, Janssen, et al., 2018a). Potential risks include inaccuracy, sensitivity, privacy, inconsistency, and data misuse (Martin et al., 2013). These risks result in governments' reluctance to open their data (Kucera & Chlapek, 2014; Zuiderwijk & Janssen, 2015). In addition, two other reasons why governments and data providers tend not to open their data are: (1) opening public and private data are a comprehensive insight that may also be able to meet disadvantages like the inappropriate interpretation of the data (Zuiderwijk & Janssen, 2013b), and (2) a mistake in translating data or misuse of the data can endanger the reputation of data providers (Barry & Bannister, 2014).

Disadvantages also include the costs of opening. Governments sometimes need considerable effort to investigate and analyse the opening data's cause and effect (Davies, 2010; Kucera & Chlapek, 2014) in order to open the data. This can be an expensive process.

Benefits and disadvantages might be related. *Causality* refers to an event or action of the disadvantages in opening data that induces something else to occur (Davies, 2010; Yang & Kankanhalli, 2013). Effect means an event or action in releasing a dataset due to another event or activity (Martin et al., 2013). For example, because of the inappropriate visualisation of a dataset in the government's portal information (as a cause), the public will tend to misinterpret the data as an effect. Unfortunately, at this moment, the investigation of the disadvantages in opening data and to what disadvantages the opening of data might result is not well-understood yet.

In this study, we categorised the economic impacts into five cost categories: collection, visualisation, management, suppression, protection, and dataset update. Data collection deals with gathering data on targeted attributes in an established system that enables answering relevant research questions (Sapsford & Jupp, 2006). Data visualisation refers to develop data and information from selected datasets clearly and efficiently (Aparicio & Costa, 2014). Data management comprises activities to manage and preserve datasets as valuable resources (Borghi, Abrams, Lowenberg, Simms, & Chodacki, 2018). Data suppression refers to the regular or ad-hoc removal of unwanted records from a contact dataset (Sweeney, 2002). Finally, data security means protecting datasets from destructive forces and the adverse actions of unauthorised users (Sun, Yongping, Zhang, & Zhu, 2014).

Table 3-4 Category and type of disadvantages in Open Government Data

	Disadvantages category (∑n=127)	Type of disadvantages	Brief description	Source
1	**Political and legislation** *(n=49)*	1.1 Data ownership *(n=5)*	Data ownership has legal rights and comprehensive control over a single piece of dataset elements. The inaccurate information about the data publishers' rightful owner of the datasets might ignore the datasets' acquisition and distribution policy.	(Al-Khouri, 2012; Evans, 2011; Kucera & Chlapek, 2014; A. Marshall, Brynjolfsson, & Madnick, 1995; Martin et al., 2013)
		1.2 Data liability *(n=7)*	Data liability is an issue limited to the data provider's side. Open data users might be afraid of being asked liable for damage caused by using the available data because of incorrect and improperly interpreted. Private organisations raised fear of liability if personal information was disclosed via an open data portal or misinterpreted by third parties to make strategic business decisions.	(Attard et al., 2015; Barry & Bannister, 2014; Chandler, 2007; Dulong de Rosnay & Janssen, 2014; Eckartz M, Hofman J, & Veenstra, 2014; Martin et al., 2013; Truli, 2018)
		1.3 Data license *(n=5)*	The data license selection process is precarious because it might not be easy to change the license status once the data is picked. The data license is often incompatible with many versions of the Creative Commons Attributions (CCA) license.	(Alamoudi, Mehmood, Aljudaibi, Albeshri, & Hamid Hasan, 2020; Giannopoulou, 2018; Mockus & Palmirani, 2015; Raffaghelli & Manca, 2019; Vir Singh & Phelps, 2007)

Disadvantages category (∑n=127)		Type of disadvantages	Brief description	Source
			Lack of attention and resources of the government organisations to check the legal protection status or license. The data license usage is often unclear with regards to metadata provenance and attributes standardisation.	
		1.4 Intellectual property rights *(n=5)*	The government should provide technological boundaries to make a better understanding of intellectual property rights use. Although open government data may be a great source in the future, it is not directly warranty covered by the intellectual property legal systems.	(Andanda, 2019; Borgesius, Frederik, van Eechoud, & Gray, 2015; Bradley, 2014; Lundqvist, 2016; Mitra-Kahn , Johnson , Man, & Meehan, 2016)
		1.5 Risk-averse culture *(n=6)*	Government institutions and external organisations with limited resources and weak links performance tend to have a risk-averse culture. This risk-averse culture can affect the decisions to open more of the datasets to the public. The decision-makers tend to be afraid of whether to open or not to open the datasets.	(Alonso, 2011; Carter, 2012; M. Janssen et al., 2012; M.-H. Lee, 2019; Martin et al., 2013; Zuiderwijk & Janssen, 2015)
		1.6 Privacy violation *(n=7)*	Open government data can serve many benefits like increasing	(Ali-Eldin et al., 2017; D. Chen & Zhao, 2012; B. Green, 2017; Korba et

	Disadvantages category (∑n=127)	Type of disadvantages	Brief description	Source
			transparency, enhancing government services, and stimulating citizen engagement. Simultaneously, government datasets might include personally identifiable data resulting in privacy issues.	al., 2008; Lundqvist, 2016; Scasa, 2014; Sweeney, 2002)
		1.7 Data access permission *(n=8)*	Data access permission refers to a data access control that open data users should apply to the data owner. For some reason, data providers close the data access because of the open data decisions' undefined objectives.	(Alonso, 2011; Arzberger et al., 2004; Carter, 2012; M. Janssen et al., 2012; M.-H. Lee, 2019; Martin et al., 2013; Yannoukakou & Araka, 2014; Zuiderwijk & Janssen, 2015)
		1.8 Data-sharing dispute *(n=6)*	Opening data to the public sometimes disputes a data-sharing protocol because the data providers do not provide documentary evidence showing the data's property rights.	(Andanda, 2019; Borgesius et al., 2015; Bradley, 2014; Lundqvist, 2016; Mitra-Kahn et al., 2016; Okediji, 2014)
2	**Technology** *(n=30)*	2.1 Data incompleteness *(n=6)*	Opening incomplete data can create a misunderstanding about the meaning of the data. The caused elements of this category are (a) the anonymity of the data source, (b) inappropriate aliases formula, and (c) mismatch of the attribute	(Amit & Larson, 1990; Kucera & Chlapek, 2014; Martin et al., 2013; Okediji, 2014; Walter, 2001; Yannoukakou & Araka, 2014)

	Disadvantages category (∑n=127)	Type of disadvantages	Brief description	Source
			relationships. This situation is also possibly influenced data quality and result in data misinterpretation.	
		2.2 Data inaccuracy *(n=5)*	Data inaccuracy can occur when data providers release their data. Some of the causes of data inaccuracy include: (a) data entry mistake by the users or data operators, (b) flawed data entry process, (c) the null problem with the value of the data, and (d) deliberate error when the users enter an ungodly amount of the data. This category can affect the quality of the data.	(D. Chen & Zhao, 2012; Dekkers, Loutas, Keyzer, & Goedertier, 2014; Kucera & Chlapek, 2014; Kučera et al., 2013; Martin et al., 2013)
		2.3 Data overlapping *(n=2)*	Datasets might contain overlapping collections of data. More datasets on various government portals might include data on a similar theme or subject. If these datasets are inconsistent users, they might get confused.	(Barry & Bannister, 2014; Kucera & Chlapek, 2014)
		2.4 Data disintegration *(n=4)*	The multi-format datasets will cause difficulties in performing the data synchronization process and consuming reassessing a dataset. Furthermore, data providers need to	(Barnickel et al., 2012); (Amit & Larson, 1990); (Martin et al., 2013); (M. Janssen et al., 2012);

	Disadvantages category (∑n=127)	Type of disadvantages	Brief description	Source
			take into account the problems from synchronization and heterogeneous datasets.	
		2.5 Low data quality (*n=5*)	Open data initiatives become useful and useable when the users can understand how to use and manipulate the released datasets. However, poor data quality can reduce the traffic to use and create inefficiency. Thus, the low data quality impacts might include user dissatisfaction, potentially increase operation costs, and a less productive decision-making process.	(Gao et al., 2016; Kučera et al., 2013; Talha, Abou El Kalam, & Elmarzouqi, 2019; Vetrò et al., 2016; Zuiderwijk, Janssen, et al., 2014)
		2.6 Data complexity (*n=8)*	Although there are many benefits of opening data to the public, the open government data movement may face complex situations. The complexities include the volume of the data, the structure of data in many relationships, and the variety of data with diverse architecture and values.	(Alonso, 2011; Barry & Bannister, 2014; Conradie & Choenni, 2014; M. Janssen et al., 2012; Jetzek, 2016; Martin et al., 2013; Toots et al., 2017; H.-J. Wang & Lo, 2016)
3	**Social** *(n=34)*	3.1 Data sensitivity *(n=6)*	Releasing data can include sensitive attributes. The users can analyse personal identity elements like full	(Barry & Bannister, 2014; Kulk & van Loenen, 2012; Martin et al., 2013; Parycek et al., 2014; Tran &

Disadvantages category (∑n=127)	Type of disadvantages	Brief description	Source
		name, date of birth, address, and phone number. This category can influence data privacy and data violation.	Scholtes, 2015; Zuiderwijk & Janssen, 2015)
	3.2 Data personal identifiable *(n=8)*	The emerging of open government data initiatives may also increase the number of data breaches that contain entities and personal identity. As a result, unauthorised users can use and analyse relevant data, such as social security numbers, passports, etc.	(Barry & Bannister, 2014; Kulk & van Loenen, 2012; Luthfi & Janssen, 2017; Luthfi, Janssen, et al., 2018a; Martin et al., 2013; Parycek et al., 2014; Tran & Scholtes, 2015; Zuiderwijk & Janssen, 2015)
	3.3 Data misuse *(n=6)*	Data disclosure can make personal or individual data identifiable by combining several datasets. Some cause misuses of the data are: (a) discredit personal profile, (b) access as unauthorised users, and (c) diminish the government's or company's reputation. This situation was influencing data privacy.	(Amit & Larson, 1990; Kucera & Chlapek, 2014; Kučera et al., 2013; Martin et al., 2013; Walter, 2001; Yannoukakou & Araka, 2014)
	3.4 Data fraud *(n=8)*	Opening data to the public may also potentially be fraud by the expert users. They can use the released datasets for several illegal actions such as detecting financial	(Barry & Bannister, 2014; Kulk & van Loenen, 2012; Luthfi & Janssen, 2017; Luthfi, Janssen, et al., 2018a; Martin et al., 2013; Parycek et al.,

	Disadvantages category (∑n=127)	Type of disadvantages	Brief description	Source
			transactions, scamming internet shopping, and filing an insurance claim.	2014; Tran & Scholtes, 2015; Zuiderwijk & Janssen, 2015)
		3.5 Data misinterpretation *(n=6)*	Publishing data by governments or companies is possible to drive a misinterpretation of the data. The causes factors of this category are: (a) insufficient domain expertise, (b) essential variables are omitted, (c) inappropriate data visualisation, and (d) error of attribute correlation. The effect of this risk category is influencing the data quality and data incompleteness.	(Amit & Larson, 1990; Barnickel et al., 2012; Barry & Bannister, 2014; M. Janssen et al., 2012; Kucera & Chlapek, 2014; Uhlir, 2009)
4	**Economic** *(n=14)*	4.1 Cost for collecting data *(n=3)*	Data collection is the process of gathering and measuring data in a sequential or systematic approach. Data collectors should follow a formal plan for data collection to ensure that the data they elicit has a precise definition, is well structured, and is accurate. Some potential costs related to this category refer to conducting a survey and investigating various data providers' data.	(Sapsford & Jupp, 2006) (Sapsford & Jupp, 2006; Ubaldi, 2013)

	Disadvantages category (∑n=127)	Type of disadvantages	Brief description	Source
		4.2 Cost for visualising data *(n=2)*	Data visualisation refers to develop data and information from selected datasets clearly and efficiently. These can help users analyse and reason about more accessible, understandable, and usable data and evidence. The potential cost to conduct this category might be developing a quantitative-based user interface of the open data portal.	(Aparicio & Costa, 2014; Xyntarakis & Antoniou, 2019)
		4.3 Cost for managing data *(n=2)*	Data management comprises activities to manage and preserve datasets as valuable resources. Some potential costs related to this category might be possible from the formatting and organising dataset.	(Borghi et al., 2018; Burwell, VanRoekel, Park, & Mancini, 2013)
		4.4 Cost for suppression data *(n=2)*	Data suppression refers to the regular removal of any unintended and anomaly records from a contacted dataset. The possible investment revenue stream from this category might reduce inaccurate data and provide various intelligent approaches to the dataset's treatment.	(S. Kim & Chung, 2019; Sweeney, 2002)

	Disadvantages category (∑n=127)	Type of disadvantages	Brief description	Source
		4.5 Cost for protecting data *(n=2)*	Data security means protecting datasets from destructive forces and the unwanted actions of unauthorised users. This category's potential costs are derived from data protection programs like data encryption, data backup, data masks, and erasure.	(Aldossary & Allen, 2016; Sun et al., 2014)
		4.6 Cost for updating data *(n=3)*	In open government data initiatives, it is essential to keep the released data updates. The costs for updating datasets include providing cloud storage and fee for administrative officers or operators.	(Borghi et al., 2018; Burwell et al., 2013; Luthfi et al., 2019)

3.3.4 Elements of decision-making support

Decision-support systems (DSS) emerged as a new domain for the first time in the '70s. They brought together the system's terminology to support managerial decisions (Filip, Zamfirescu, & Ciurea, 2017). For one reason, the DSS concept was prevented by the idealised vision of over the precognitive man-computer systems, enabling man and computers to gain cooperation on making decisions and managing various complex situations (Licklider, 1960).

The DSS can be divided into six main types (Bonczek, Holsapple, & Whinston, 1980; Turban, Aronson, & Liang, 2007). First, model-driven DSS, which is based on quantitative models. This first model provides the most elementary functionality of the decision-making process. The functionality includes three main parts, namely, planning, scheduling, and management. The decision-makers tend to use this model to design a simple decision-making process within a short period of time. Second, data-driven DSS emphasised the access and manipulation of data modified to specific works using general decision-making tools. This model uses the elementary functionality of decision-making processes like planning, scheduling, and management for supporting decisions in a range of specific cases and situations. Third, communication-driven DSS uses communication and network technologies to support and facilitate the decision-making process. This model helps the collaboration and communication aspects and possible to use various tools, including computers and communication supporting tools. Fourth, document-driven DSS uses large document databases that store several file formats like word processing documents, images, videos, and other multimedia platforms. Fifth, knowledge-driven DSS introduces human-computer systems that come up with problem-solving expertise. This model combines the artificial intelligence field and human cognitive capabilities to provide suggestions to decision-makers. Sixth, web-based DSS is the latest and the most sophisticated decision support system model that extends its capacities and abilities using Internet-based technologies.

There is no consensus on a DSS's capabilities because of a shortfall of general agreement about the DSS domain (Lamy, Ellini, Nobécourt, Venot, & Zucker, 2010;

Omidvar & Bordbar, 2013). Nevertheless, most of the DSS has two main features that could help the decision-makers. First, decision-makers at all levels of an organisation, individual, group of users, mainly semi-structured and non-structured situation, with actual data and supporting human rationalities (Lamy et al., 2010; Omidvar & Bordbar, 2013; Turban et al., 2007). Second, a DSS can improve the decision-makers insights and knowledge from a specific type of case study (Turban et al., 2007). Thus, a DSS leads to new demands, achieves the organisation's objectives and agendas, and provides a knowledge-based repository of a specific domain (Omidvar & Bordbar, 2013; Sauter, 2010).

The essential purpose of a DSS is to provide useful information to decision-makers for making a decision (Bonczek et al., 1980; Power, 2002; Turban et al., 2007). Therefore, our DSS should collect relevant information from the knowledge repositories, analyse it using an appropriate method, and present it to the decision-makers and other related stakeholders (Filip et al., 2017; Sauter, 2010; Turban et al., 2007). At this moment, there is no universally accepted taxonomy of the decision-support systems because different authors provide and propose different categories and elements (Ahmad Mir & Quadri, 2009; Power, 2002). Every DSS does not well-suited into a single category but is a combination of several references (A. Marshall et al., 1995; Omidvar & Bordbar, 2013). Therefore, we define the theory of DSS's elements by investigating a systematic literature review. We used the Elsevier Scopus database to cover both topical and non-topical articles. Eligible papers represented the key construct of "decision support systems".

Table 3-5 summarises the architectural elements of DSS provided by the different authors and sources.

Table 3-5 Overview of DSS elements from the literature

	Source	Elements of DSS	Description of elements
1	(Sprague & Carlson, 1982)	a) Database Management System (DBMS) b) Model Base Management system (MBMS) c) Dialogue Generation and Management System	a) Stores information b) Integrates models c) Provides user interfaces to manage system
2	Haettenschwiler (1999)	a) User authentication b) Decision context c) Target system d) Knowledge-based	a) Participates in different roles or functions in the data management process b) Specifically defined decision rules c) Describes the majority of the preferences d) External data sources, knowledge databases, working databases, data warehouse, metadata bases, models, methods, integrates search engines to responding system
3	(Marakas, 1999)	a) Database Management System b) Memory Management System c) Knowledge Engine d) User Interface e) User	a) Stores, manages and provides access to the data b) Organises memory efficiently c) Inference procedure or control structure for utilising the knowledge d) Allows a user to interact with the system e) One who uses the system
4	(Haag, Cummings, & Dawkins, 2000)	a) Database Management System (DBMS) b) Model Base Management system (MBMS) c) Dialogue Generation and Management System	a) Stores information (that can be further subdivided into the organisation's traditional data repository, from an external source such as the Internet or the experience of the individual user)

	Source	Elements of DSS	Description of elements
			b) Using various kinds of models, it handles the representation of events, facts or situations c) Integrates models and provides a user interface
5	(Power, 2002)	a) The user interactive b) The database interactive c) The model and analytical tool d) The DSS architect and network	a) Interacts with the user over a command line b) Interacts with a single or a group of users using a database for heuristics c) Model designed for analysis d) Interacts with the other DSS or database server

3.4 Taxonomy and elements of the DSOD

This study developed a taxonomy of the advantages and disadvantages of opening data from the systematic literature review. Based on the systematic literature review that was carried out in sections 3.3.2 and 3.3.3, we clustered the advantages and disadvantages of opening data at the same level into three main categories: political and legislation, technology, social, and economic elements. In the advantages cluster, the political and legislation is viewed in eight sub-categories: improved transparency, enhanced accountability, political awareness, improved policy-making, increased reputation, improved government services, data sharing agreement, and evidence-based policy-making. In the disadvantages cluster, the political and legislation cluster consists of eight categories: data ownership, data liability, data license, intellectual property rights, risk-averse culture, privacy violation, data access permission, and data-sharing dispute.

The main category's technology advantages consist of seven sub-categories: linked open data, data optimisation, data exploration, data discovery, data validation, data combination, and data machine-readable. The disadvantages of the technology aspect consist of six sub-categories, e.g., incompleteness, inaccuracy, overlapping, disintegration, low data quality, and complexity of the data.

In the main social category, the advantages of disclosing data are classified into eight sub-categories: improving citizen engagement, data reusability, better decision-making, combating corruption, improving public trust, increasing public satisfaction, data availability, and encouraging public literacy. The disadvantages of opening data from the social aspect consist of five categories: data sensitivity, personally identifiable data, data misuse, data fraud, and data misinterpretation.

Moreover, in the main economic category, the advantages of opening data can contribute to the five sectors: knowledge-economy growth, innovation, increased efficiency, new business opportunities, and new job creation. Simultaneously, opening data to the public can create costs from the economic perspective, such as collecting data, visualising data, managing data, suppressing data, protecting data, and updating data. Figure 3-2 presents the taxonomy of the advantages and disadvantages of opening government data.

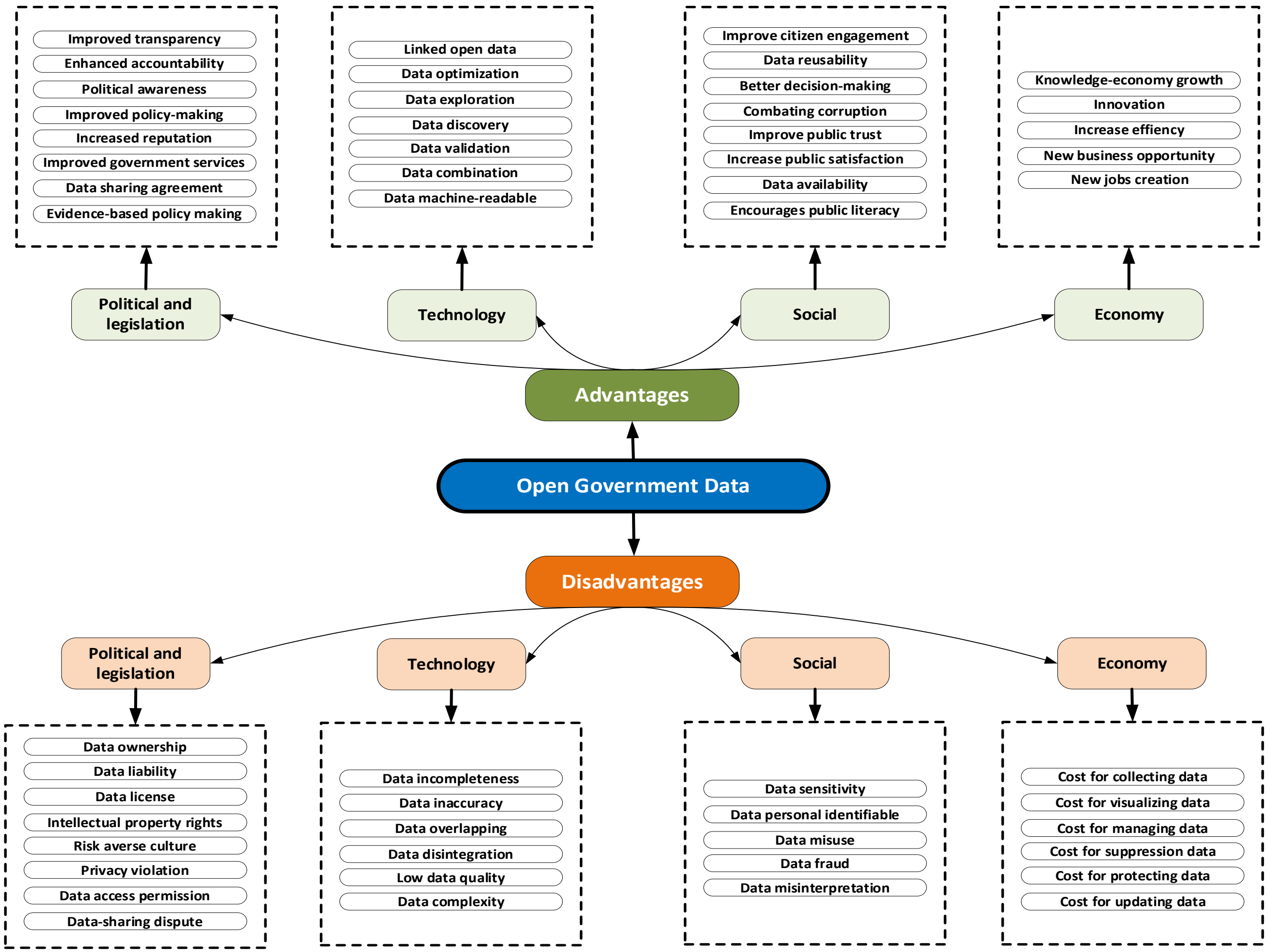


Figure 3-2 Taxonomy developed from the literature review

In addition, this study also provides the elements of the decision-making process to open data. Based on the systematic literature review in section 3.3.4, we found eight elements of the DSOD in this study. First, the Database Management Systems (DBMS) refers to the stores of information that can be further subdivided into the organisation's traditional data repository, from an external source such as the Internet or the individual user's experience. Second, the Model Base Management system (MBMS) refers to the use of various kinds of models. It handles the representation of events, facts, or situations. Third, Dialogue Generation and Management System (DGMS) refers to integrating models and provides a user interface. Fourth, the user interface refers to the inference procedure or control structure for utilising the knowledge. Fifth, user authentication represents participating in different roles or functions in the data management process. Sixth, decision context refers to the specifically defined decision rules. Seventh, the knowledge-based refers to the external data sources, knowledge databases, working databases, data warehouse, metadata bases, models, methods, integrates search engines into the responding system. Eighth, the model and analytical tool refer to the model designed for analysis. Table 3-6 shows the decision-making support elements that we can use as the fundamental components to develop the DSOD prototype in this research.

Table 3-6 Elements of DSS used in this research

	Elements of DSS derived from literature	**Elements of DSS use in this study after combining and merging similar term**	**Description of DSS elements**
1	a) Database Management System (DBMS) b) Model Base Management system (MBMS) c) Dialogue Generation and Management System	*a) Database Management System (DBMS)* *b) Model Base Management system (MBMS)* *c) Dialogue Generation and*	a) Stores information (that can be further subdivided into the organisation's traditional data repository, from an external source such as the Internet or the experience of the individual user) b) Using various kinds of models, it handles the representation of events, facts, or situations c) Integrates models and provides a user interface
2	a) User authentication		

	Elements of DSS derived from literature	Elements of DSS use in this study after combining and merging similar term	Description of DSS elements
	b) Decision context c) Target system d) Knowledge-based	*Management System (DGMS)* *d) User interface* *e) User authentication* *f) Decision context* *g) Knowledge-based* *h) The model and analytical tool*	d) Inference procedure or control structure for utilising the knowledge e) Participates in different roles or functions in the data management process f) Specifically defined decision rules g) External data sources, knowledge databases, working databases, data warehouse, metadata bases, models, methods, integrates search engines to responding system h) Model designed for analysis
3	a) Database Management System b) Memory Management System c) Knowledge Engine d) User Interface e) User		
4	a) Database Management System (DBMS) b) Model Base Management system (MBMS) c) Dialogue Generation and Management System		
5	a) The user interactive b) The database interactive c) The model and analytical tool d) The DSS architect and network		

3.5 Conclusions

In this chapter, a detailed and systematic literature review was presented to answer the first research question (RQ#1), namely "*What are the advantages and disadvantages of opening data?*", and the second research question (RQ#2), "*What are the elements of the decision-making support for opening data?*". From the systematic literature review, a taxonomy of the advantages and disadvantages of opening data was developed. This taxonomy contributes to the scientific knowledge in open government data, and RQ#1 was answered. Based on the systematic literature review that we carried out in sections 3.3.2 and 3.3.3, we categorised the advantages and disadvantages of opening data at the same level into four categories: political and legislation, technology, social, and economic aspects.

Furthermore, this chapter also provides the elements of the decision-making process to open data. Based on the systematic literature review in section 3.3.4, we found that eight elements of relevance for the DSOD in this study. The eight elements of the DSS include (1) Database Management System (DBMS), (2) Model Base Management system (MBMS), (3) Dialogue Generation and Management System (DGMS), (4) User interface, (5) User authentication, (6) Decision context, (7), Knowledge-based, and (8) The model and analytical tool. Based on these eight elements of the DSS found through literature studies, we then translated the DSS's main requirements according to the DSOD needs in Chapter 4.

Chapter 4
Decision-making support functionalities

This chapter presents the functionalities of decision-making support to open data (DSOD). This chapter's main objective is to answer the third research question (RQ#3): What are the functionalities of a prototype? To address the research question, we followed a prototyping developing approach. This chapter used the literature review of decision-making support elements in the previous part (Chapter 3) to define and describe the decision-making support. Five functionalities of decision-making support contribute to developing a DSOD prototype. The functionalities include retrieving and decomposing the dataset, evaluating the dataset, assessing and weighing the dataset, providing decision alternatives, and providing recommendations.

This chapter is structured using five sections. First, we introduced the need to develop decision-making support to open data. Second, we provided the phases of designing the DSOD, which consisted of the Intelligence phase, design phase, and choice phase. Third, we defined the functionality of the DSOD elements provided in Chapter 3. Four, we translated the DSS elements found in chapter 3 to DSOD functionalities. In the last sub-section, we use derived and presented the detailed DSOD steps.

Parts of this chapter have been published in (Luthfi & Janssen, 2017, 2019b; Luthfi, Janssen, et al., 2018b, 2020).

4.1 Decision-making support for opening data

The government expects to provide open access to the datasets for the public. The opening of data should result in the accomplishing of public values, like transparency and accountability, but at the same time, other public values like privacy should be ensured (Zuiderwijk & Janssen, 2015). Although the opening might yield benefits, they might also encounter risks (Conradie & Choenni, 2014; Zuiderwijk & Janssen, 2017). Possible disadvantages are the misuse of data or information to benefit individuals, groups, or even politicians (Kulk & van Loenen, 2012). Another frequently mentioned reason is privacy that can lead to inappropriate interpretation by the public and hinder contributing to the spirit of open data (M. Janssen & Van den Hoven, 2015; Zuiderwijk & Janssen, 2015).

Although many datasets have been opened (Grimmelikhujsen & Meijer, 2014; Kulk & van Loenen, 2012; Meijer & Thaens, 2009) a substantial number of datasets is still closed (Zuiderwijk & Janssen, 2015). Several reasons can explain this situation, including the reluctance of some organisational entities to release datasets for several reasons, including the complexity of implementing systems (Barry & Bannister, 2014; Martin et al., 2013; Veenstra & Broek, 2013), high human skills and well-educated staff (Albano & Reinhard, 2014; Puron-Cid et al., 2012; Yannoukakou & Araka, 2014), and the readiness related to the infrastructure, hardware, software, or financial resources (Gurstein, 2011). The disadvantages of opening data include mistakes in data and potential misuse of data, which can endanger data providers' reputations (Barry & Bannister, 2014). One particular disadvantage is that opening a personal data system violates the data protection act concerning data privacy (Kulk & van Loenen, 2012). When datasets are linked together, at the same time, the risk of privacy violations will increase (Bertot et al., 2010). Overall there is a lack of insight into the potential disadvantages and advantages of open data (Manyika et al., 2013).

In the literature study, we found that there are various models for making decisions to open data. The five systematic models contributing to the open data domain were identified: (1) *Trade-off the risks values* (Zuiderwijk & Janssen, 2015). This model provides structured steps for analysing the benefits and risks of disclosing data. (2) *Decision-support framework* (Buda et al., 2015). This model provided a prototype that was based on the insight of open data ecosystems. (3) *Multiple Criteria for decision-making* (Luthfi, Janssen, et al., 2018a). This model used a fuzziness theory to analyse uncertainty problems and provide decision alternatives. (4) *Costs and benefits of opening data* (Luthfi & Janssen, 2019b). This model was developed based on the DTA method. This model is used to estimate the potential advantages and disadvantages of releasing data. (5) *Interactive decision-making process* (Luthfi &

Janssen, 2017; Luthfi, Janssen, et al., 2018a). This model proposed a BbN method to construct the causal relationships of the decision-making process to open data in the case of health patient records. This model contributes to the perspective of how to examine the risks and benefits of opening data by providing sequential iteration process. The model uses a suppression technique like k-anonymity to anonymise such sensitive attributes.

4.2 Designing DSOD phases

Decision-making support (DMS), on the other hand, is a specific form of information processing, which is presenting an action plan under particular circumstances (Simon, 1979). Herbert Simon introduces three primary phases to develop decision-making support, as described in Chapter 3. In this study, the development of a decision-making support system will follow Simon's process model of the decision-making process and its elements, as presented in Figure 4-1.

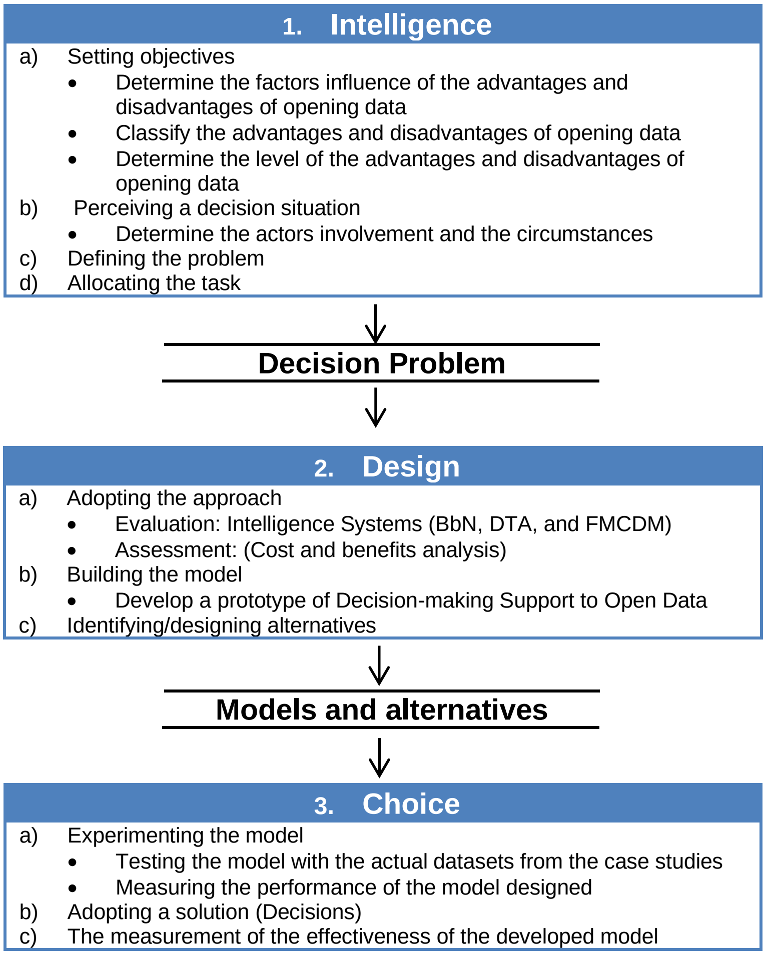


Figure 4-1 Decision-making support for opening data process

1) **Intelligence phase**. A DSS development project should start by obtaining a clear understanding of the decisions to be made with the proposed system's help (Turban et al., 2007). The intelligence phase's term refers to collecting information without knowing the decision's outcome (Mallach, 1994). The intelligence phase may involve activities (Turban et al., 2007), such as (a) setting objectives and observe the problems, (b) develop statements of the problem and perceiving a

decision situation, (c) acquiring information relevant to the decision, and (d) allocating the task.

2) **Design phase**. The design phase involves systematic and well-structured research to determine alternatives or available options (Mallach, 1994). Several activities will be carried out, including (a) identify variables and criteria for the decision, (b) specify relationships between variables, (c) Identify controllable and uncontrollable events, (d) develop a decision-model that can be used to evaluate alternatives and (e) evaluation of a variety of possible solutions to the problem statements.
3) **Choice phase**. This final step is called the decision, which presents the way to release it for implementation. In the choice phase, alternatives are searched, evaluated, and one chosen as a recommended solution (Simon, 1979). The selected decision is to be carried out and only if the recommended solution is successfully employed and the problem is solved (Turban et al., 2007). Three activities will be carried out, namely (a) experimenting with the model, (b) adopting a solution, and (c) measurement of the effectiveness of the developed model.

4.3 Elements of the DSOD

In this chapter, we translated the eight elements found in Chapter 3 to develop the DSOD prototype. Based on the systematic review in Chapter 3 (Section 3.3.4), the main elements were structured into eight parts and are detailed for the DSOD prototype in this section. First, the Database Management Systems (DBMS) refers to the stores of information that can be further subdivided into the organisation's traditional data repository, from an external source such as the Internet or the individual user's experience. Second, the Model Base Management system (MBMS) refers to the use of various kinds of models. It handles the representation of events, facts, or situations. Third, Dialogue Generation and Management System (DGMS) integrates models and provides a user interface. Fourth, the user interface refers to the inference procedure or control structure for utilising the knowledge. Fifth, user authentication refers to the participation of different roles or functions in the data management process. Sixth, decision context refers to the precisely defined decision rules on accepting or rejecting alternative decisions. Seventh, the knowledge-based refers to the external data sources, knowledge databases, working databases, data warehouse, metadata bases, models, methods, integrates search engines into the responding system. Eighth, the model and analytical tool refer to the model designed for analysis. Table 4-1 shows the decision-making support elements that we can use as the fundamental components to develop the DSOD prototype in this research. Table 4.1 presents the eight DSS elements derived from the literature and generally used for defining and designing decision-making support.

Table 4-1 Elements of DSS used in this study

	Elements of DSS from literature	Functionality
1	Database Management System (DBMS)	DBMS is an intermediary system that is used as a good liaison between the user and the database. The DBMS aims to function as a tool for organising well-structured corporate data sources (Mallach, 1994; Sprague & Carlson, 1982).
2	Model Base Management system (MBMS)	MBMS is an interactive system consisting of a user dialogue system, a model processor, and a data management system, which helps decision-makers use quantitative data and models to solve semi-structured problems (Koutsoukis, Dominguez-Ballesteros, Lucas, & Mitra, 2000; Sprague & Carlson, 1982).
3	Dialogue Generation and Management System (DGMS)	The creation of a dialogue generation and management system (DGMS) aims to increase the tendency and ability of system users or stakeholders to gain the most from a DSS. Since most DSS uses are optional, the decision-maker must be motivated to use a DSS or open up a great chance of remaining unused (Power, 2002; Turban et al., 2007).
4	User Interface	The user interface is what the user sees and uses when interacting with the DSS. Most users do not want to learn a more technically oriented interface. For non-technical users, a suitable DSS user interface design is the essential determinant of a successful decision support implementation (Bonczek et al., 1980).
5	User authentication	User authentication is the act of verifying statements, such as the identity of a DSS system

	Elements of DSS from literature	Functionality
		user. Utilising user authentication enables the DSS system to verify a person's identity connected to a network resource (Koutsoukis et al., 2000).
6	Decision context	Designing a decision context aims to define what decisions are made and why. Does it relate to other decisions made or previously anticipated? How to rank decision risk, and why? How will the information be used in decisions made? (Ahmad Mir & Quadri, 2009)
7	Knowledge-based	Knowledge-based decisions are used to make effective and strategic decisions by establishing the thought process and rationale behind decisions. What factors influence the decision, and how are the strategies for keeping the objectives of a DSS? (Turban et al., 2007)
8	The model and analytical tool	Models and analytical tools in DSS are used to analyse complex situations related to decision-making. This model allows scientific and experimental approaches with different strategies to find the right method for an organisation or DSS users (Omidvar & Bordbar, 2013; Sprague & Carlson, 1982).

4.4 Translation of the DSS elements to DSOD functionalities

Based on these eight elements of the DSS found through literature studies, we translated the DSS's main requirements according to the DSOD needs. This translation process aims to (1) redefine each element according to the DSOD development stage, (2) explain in more detail the role of each translation element during DSOD development, and (3) as a bridge to translate between literature studies on DSS. Figure 4-2 shows the translation process from DSS elements derived from the literature studies to the DSOD functionalities.

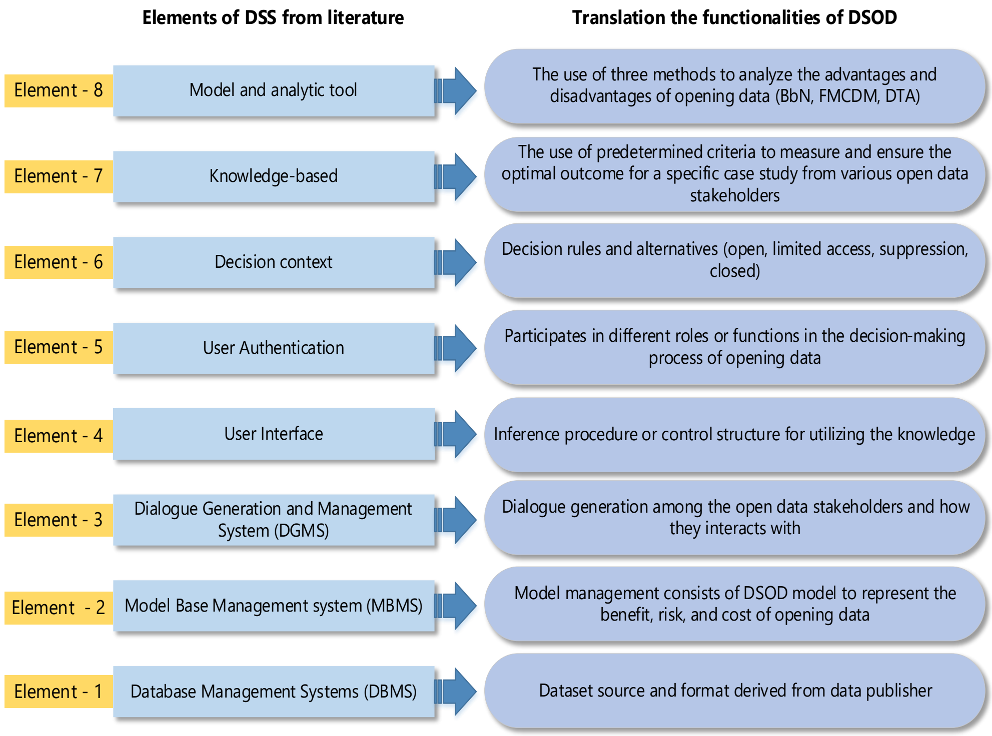


Figure 4-2 Translation of the DSOD functionalities

The translation of the DSS elements to DSOD functionalities is explained in more detail hereafter.

In the first element, Database Management Systems (DBMS) represent the need for data providers to set the dataset source, including several technical approaches, such as extracting, transforming, and loading (ETL) a dataset. The DBMS enables users to use and access dataset together using multi-database platforms. Several dataset formats could be used in this context, such as CSV, JSON, XML, AACDB, and other machine-readable formats.

As identified for the second element, the Model Base Management System (MBMS) represents an interactive system consisting of a user dialogue system, which helps decision-makers analyse datasets to solve semi-structured problems. In the context of the DSOD functionality, MMBS can determine the advantages and disadvantages of opening the datasets to help decision-makers, civil servants, and administrative officers.

In the third element, the Dialogue Generation and Management System (DGMS) aims to increase the tendency and ability of system users or stakeholders to gain the most from a DSS. The stakeholder's involvement in the decision-making process to open data might have different roles and objectives. The various concerns, like transparency and privacy, should be represented and discussed. Our taxonomy will help to ensure that all benefits and disadvantages will be taken into account. Therefore, we use the DGMS to generate a dialogue generation among the open data stakeholders and how they interact with each other.

The user interface aims to provide an interface about how the users interact with the DSS as positioned in the fourth element. The user interface should ensure that all information for making a decision can be accessed. In the functionality of DSOD, we design a user interface by considering technical and non-technical barriers. Therefore, we develop a DSOD prototype to help decision-makers, data analytics, and civil servants use the DSOD conveniently.

As defined in the fifth element of DSS, user authentication enables the DSS system to verify a person's identity. The user authentication should ensure that DSS gives permissions to the users to access a resource. In the functionality of the DSOD, we developed a user authentication to verify statements, such as the identity of a DSS system user, roles, and their privilege in using the DSOD.

The sixth element of the DSS, namely decision context, is to define what decisions are made and why. Does it relate to other decisions made or previously anticipated? How to rank decision risk, and why? How will the information be used in decisions made? The decision context may enable the re-use of past decisions made. When similar datasets are disclosed, the current data could be re-used and re-analysed. Therefore, it might be more efficient because no new process needs, except when the circumstances are altered. In the functionality of the DSOD, we used decision context to create decision rules and alternatives, namely open, limited access, suppression, and closed the dataset.

The seventh element, knowledge-based, aims to make effective and strategic decisions by establishing the thought process and rationale behind decisions. What factors influence the decision, and how are the strategies for keeping the objectives of a DSS? In the translation of the DSOD functionality, knowledge-based defines predetermined criteria to measure and ensure the optimal outcome for a specific case study from various open data stakeholders. Similar to the decision context in the sixth element, if a dataset has already opened, the current knowledge could be re-used for making new decisions.

The last and eighth element, namely the model and the analytic tool, is used to analyse complex situations related to decision-making. This model allows scientific and experimental approaches with different strategies to find the right method for an organisation or DSS users.

After we translated the DSS's needs and elements, the next step was to build a DSOD conceptual model. To do this, we combined the elements of a DSS based on literature studies with facts and decision-making models from early studies in several departments in government organisations in Indonesia. The process steps for developing the DSOD conceptual model can be explained as follows:

First, we build the first step called retrieving and decomposing the dataset. This step is a combination of the translation results from element 1 (DBMS), element 2 (MBMS), element 4 (user interface), and element 5 (user authentication). At this step, DSOD emphasises the aspects of how DSOD initialises the process of retrieving and parsing the dataset, including extracting data, creating metadata, and setting user authentication.

Second, we design a second step called the evaluation stage. This step is a combination of the translation results from element 3 (DGMS), element 7 (Knowledge base), and element 8 (Model and analytic tool). At this stage, DSOD is designed to evaluate the potential disadvantages of opening the data by considering the aspects of the benefits at the same time. In this step, several data analysis methods such as BBN, FMCDM, and DTA are used. Meanwhile, a knowledge base of decision-makers, politicians, and other stakeholders is needed to quantify the disadvantage categories and benefits of data disclosure.

Third, we prepared an assessment model from the third step's evaluation results after conducting the evaluation. In this step, we performed a translation of element 4 (user interface) and element 7 (knowledge base). At this step, the assessment is carried out using a pairwise comparison matrix to weigh the advantages and disadvantages of opening the data.

Fourth, we provide several alternative decisions based on the results of the assessment in the third step. This alternative decision was made through a literature study and adopting a decision model obtained when conducting initial studies in several government institutions in Indonesia. Some alternative decisions are open, limited access, suppression, and closed dataset.

Finally, we provide recommendations for decisions based on the assessment results and alternative decisions taken in the fifth step. This recommendation is a technical step on how to treat a decision result on the dataset's status. If the decision's outcome is to give special treatment to the dataset, then DSOD will provide technical steps, such as anonymisation or removing some other disadvantage attributes. Figure 4.3 presents the translation of the DOSD development in this study.

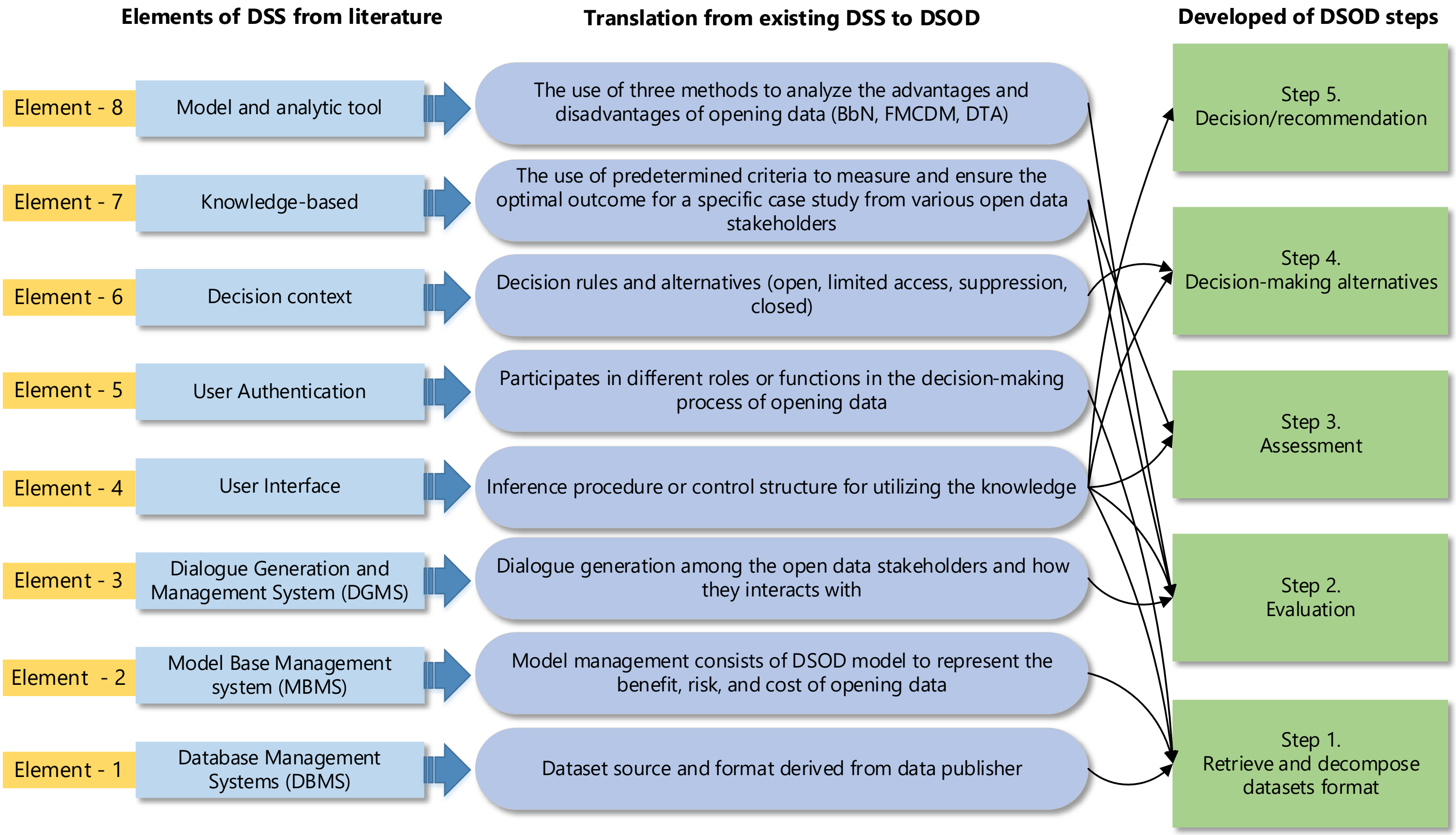


Figure 4-3 Translation of the functionalities and development to the DSOD steps

4.5 Steps of the DSOD

As a general overview, this study uses a hierarchical process that starts with selecting public datasets. Datasets can be collected from data providers, government departments, or data agencies (Ubaldi, 2013). In the next step, datasets are processed on a prototype of the Decision-making support system. In this step, a variety of methods in this research, BbN, FMCDM, and DTA, will be taken apart for the suitable determining approach in the evaluation process. The Decision-making Support System model for opening data starts with the retrieval of datasets in a data provider. There are five main steps to judge whether to open data. The steps used in this model present in Figure 4-4, as follows:

- Step 1. Retrieve and decompose datasets format. The system will carefully sort the tables in the datasets, including ensuring that all fields are intact and maintained in relation to each table. A variety of datasets structure formats can be read by the system both based on proprietary database or open platforms such as ACCDB (Microsoft), CDF (XML standard), Comma-separated Value (CSV), Database File (DBF), Digital Elevation Model (DEM), ESRI (Geo DB), JavaScript Object Notation (JSON), and others.
- Step 2. Evaluation. At this stage, the datasets that have been selected in the previous step will go through the evaluation process. The system will interpret data that translates each data value from a table to be included in two broad categories of advantages and disadvantages. Datasets are evaluated using the Intelligence System Algorithms such as DTA, BbN, or FMCDM. In the case of using BbN, there are seven stages to run the evaluation process. (See Table 4-5). This step's output is the list of advantages and disadvantages of the dataset.
- Step 3. Assessment. The previous stages' evaluation results are classification and level references to the advantages and disadvantages of datasets. This system's advantage is to provide iterative process conditions when conducting an assessment to ensure that the benefits level is higher than the

disadvantage at hand. Technically, during the assessment process, the system will combine the overall scores from the benefit and risk analyses to determine the appropriate solution for treating the dataset.

- Step 4. Decision-making alternatives. There are four possible decisions to release the datasets. *Open*: Publishing the dataset presents a low disadvantage to the individual or organisational identification, or the potential benefits of the dataset substantially outweigh the potential risks. *Limited Access*: Publishing the dataset will create a moderate disadvantage, or the potential benefits of the dataset do not outweigh the potential disadvantages. *Additional Screening*: publishing the dataset makes significant risks, and the potential benefits do not outweigh the potential disadvantage. *Closed*: Releasing the dataset generates a high or very high disadvantage to the individual or organisation and significantly outweighs the potential benefits.
- Step 5.a. Open Decision. In this step, when the datasets test results show that the datasets have a higher significant advantage condition than the disadvantage, the system will provide a reference to open the data to the public.
- Step 5.b. Non-open Decisions. Suppose the risk attribute of the datasets is still higher than the advantage (5.b.1). In that case, the system performs the iteration of the trace hold disadvantage. It returns to the evaluation in step 3 (5.b.2) until the disadvantages are reduced or removed, and the benefit increased. The possibility of non-open decisions is (Limited Access, Additional Screening, or Closed).

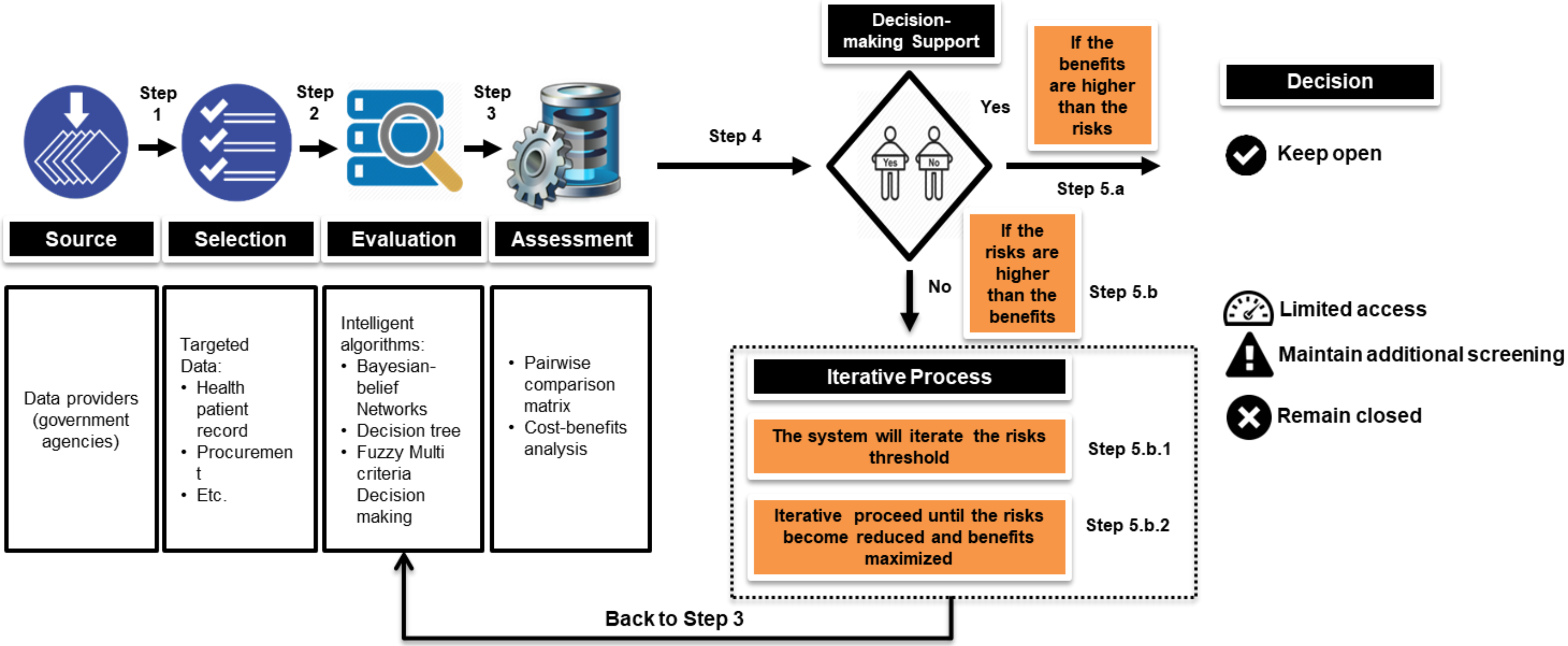


Figure 4-4 Steps of the DSOD
(Luthfi & Janssen, 2017)

Based on the steps of the decision-making process to open data developed in the previous discussion, we then used the five main sub-steps, namely (1) retrieve and decompose datasets, (2) analysing the dataset, (3) weighing the dataset, (4) decision-making alternatives, and (5) updating the status of the dataset. This part provides a detailed process by using the decision-making process elements to open data in Chapter 3.

In the domain of open government data, a DSS is proposed to help the decision-makers analyse the potential disadvantages of opening a dataset by considering the benefits at the same time. The proposed DSS supports in the decision-making process include constructing the causal relationship between advantages and disadvantages using three methods: BbN, FMCDM, and DTA.

Chapter 5
Methods for open data decision-making

In this chapter, the three methods, namely Bayesian-belief Networks (BbN), Fuzzy Multi-Criteria Decision-making (FMCDM), and Decision Tree Analysis (DTA), are used in the functionalities of the DSOD presented in Chapter 4. The methods are explained, and an illustration is given in this chapter. These methods will be used to analyse the potential advantages and disadvantages of opening data. We selected three different methods because we could not identify a single best method based on our three selection criteria. The reason for this is that various stakeholders have different roles and interests and might favour different methods. Therefore, a variety of methods was chosen.

The BbN is chosen as this method can capture the probabilistic relationship between factors and the opening data decisions by taking into account conditions and external events. The FMCDM can support decision-making by experts within a short time to work with various inputs from the different stakeholders' backgrounds. The DTA provides a mechanism for dealing with the disadvantages, benefits and estimating the possible consequence. The proposed three methods contribute to the knowledge of decision-making support in the open government data domain. Parts of this chapter have been published in (Luthfi & Janssen, 2019a; Luthfi, Janssen, et al., 2018a; Luthfi et al., 2019; Luthfi, Rehena, Janssen, & Crompvoets, 2018).

5.1 Bayesian-belief Networks

Bayesian Networks (BNs), also known as Bayesian-belief Networks (BbNs) or Belief Networks, are probabilistic graphical models that represent a set of random variables and their conditional dependencies via a directed acyclic graph (DAG) (Pearl & Russel, 2001). The conditional dependencies are the relationship between two or more events that are relevant to the third event to occur (Husmeir, 2005). For example, when new datasets are opened, the change to reidentification of already opened datasets increases (Koot, Noordende, & Laat, 2010; Rocher, M. Hendrickx, & Montjoye, 2019). Thus, the chance of re-identification is dependent on the condition that other datasets are opened. The BbN can be used to explore and display causal relationships between key factors and the system's final outcomes. Hence, it can look at the causal relationship of factors influencing the impact when data is opened. As BbN's create a causal model, they can also be used to calculate the effectiveness of interventions, such as alternative management decisions or policies, and system changes, such as those predicted for earthquake risks probabilistic (Bayraktarli & Ulfkjaer, 2005). One advantage is that the uncertainties associated with these causal relationships can also be explored simultaneously (Ben-Gal, 2008). This helps us to understand better the cause and effect of the disclosing datasets. Furthermore, the BbN's can maintain clarity by making causal beliefs explicit (Cárdenas, Halman, & Al-Jibouri, 2012) and are often used to model when the relationships cannot easily be expressed using mathematical notation (Neopolitan, 2003; Pearl & Russel, 2001).

5.1.1 Uncertainties in Bayesian-belief Networks

The BbNs originate from research into Artificial Intelligence (AI), where they were originally constructed as a formal means of analysing decision strategies under uncertain conditions (Chockalingam, Pieters, Teixeira, & Van Gelder, 2017). The BbNs are particularly useful for diverse problems of varying size and complexity, where uncertainties are inherent in the system (Chakraborty, Mengersen, Fidge, Ma, & Lassen, 2016). For example, uncertainty in our situation can refer to the development of more advanced techniques enabling the reidentification of datasets. This can influence decision-making as reidentification might be possible in the future.

Bayesian networks apply Bayes’ Theorem, also known as Bayes’ Rule or Bayes’ law (Neopolitan, 2003; Spiegelhalter, 1998). In Bayes’ theorem, a prior or unconditional probability represents the likelihood that an input parameter will be in a particular state (Neopolitan, 2003). The conditional probability calculates the likelihood of the state of a parameter given the states of input parameters affecting it (Murphy, 1998). The posterior probability is the likelihood that the parameter will be in a particular state, given the input parameters, the conditional probabilities, and the rules governing how the possibilities combine (Heckerman, 2008; Pearl & Russel, 2001). In our case, conditions help weigh the benefits and disadvantages of the selected dataset by considering uncertainties and conditional dependencies driven by new external datasets.

The "Networks" is solved when nodes have been updated using Bayes' Rules, as follows (Bøttcher & Dethlefsen, 2003):

$$P(A|B) = \frac{P\ (B|A)\ P(A)}{P(B)}$$

Where P(A) is the prior distribution of parameter A; P(A|B) is the posterior distribution, the probability of A given new data B; and P(B|A) the likelihood function, the probability of B given existing data A. For example, A could be the change of reidentification, and B could be the opening of another dataset.

Dissimilar to many other predictive modelling techniques, BbNs use the probabilistic approach rather than deterministic expressions to describe the relationships among variables (Cárdenas et al., 2012). The probability is influenced by the information about an event's possible occurrence, like that dataset enabling reidentification are likely to be opened by other organisations. Expert knowledge is accounted for in the network by applying the Bayesian probability theory (Liu, Chen, Lu, & Shen, 2012). Therefore, the BbN's theory allows subjective assessments of the probability that a particular outcome will be combined with more objective data quantifying the frequency of occurrence in determining conditional probabilistic relationships (Pearl & Russel, 2001). In addition, BbNs have several other appealing properties that make them particularly useful for data analysis and decision-making (Cárdenas et al., 2012; Robertson & Wang, 2004).

As BbNs are causal, they can be used to quantify the effectiveness of interventions, such as alternative policy-maker decisions whether to open or not to open a dataset. Notably, the uncertainties associated with causal relationships can also be explored at the same time (Murphy, 1998; Nadkarni & Shenoy, 2001). For example, a dataset containing sensitive personal attributes, such as name, date of birth, home address, and citizen identification number. Opening these sensitive attributes to the public domain, on the one hand, can result in several disadvantages, like personal identification and not complying with the GDPR. On the other hand, disclosing the attribute may provide transparency for certain stakeholders. The BbNs can construct causal relationship factors influencing the advantages and disadvantages of opening data by downward evidence propagation.

5.1.2 Structure of Bayesian-belief Networks

A Bayesian network uses probability and graph theory to construct probabilistic inference and reasoning models (Neopolitan, 2003; Pearl & Russel, 2001). It is described as a Directed Acyclic Graph (DAG) with nodes and arcs. Nodes represent variables, events or evidence (Fenton & Neil, 2012; Pearl & Russel, 2001). An arc between two nodes represents a conditional dependency between the nodes (Fenton & Neil, 2012; Pearl & Russel, 2001). Furthermore, arcs are unidirectional, and feedback loops are not accepted (Fenton & Neil, 2012; Nadkarni & Shenoy, 2001; Pearl & Russel, 2001). Therefore, it is possible to identify the parent-child relationship or the probability dependency between two nodes because of this characteristic (Robertson & Wang, 2004).

The structure of a Bayesian network can be described graphically, where variables (or nodes) are connected by unidirectional arrows (or arcs) (Pearl & Russel, 2001). A BbN is constructed as a causal structure, where node A (open dataset) affects node B (data sensitivity), which may affect node C (data misuse), as presented in Figure 5.1. In this case, A is referred to as a parent of B, with B being referred to as a child of A. B will thus be a parent of C and is sometimes referred to as an intermediate node.

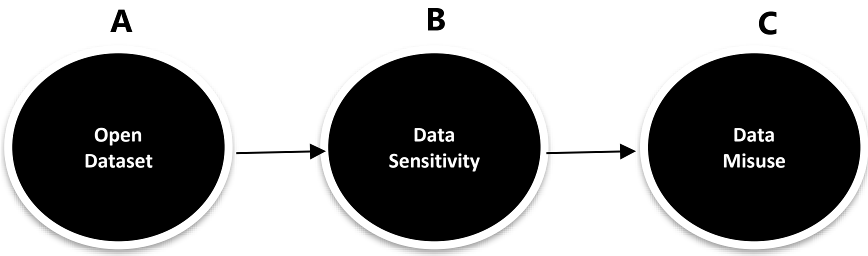


Figure 5-1 Basic causal structure of a BbN

In a BbN rule, arcs' directions cannot loopback (i.e., cycle back into the model), and the form of the structure is a DAG (Chakraborty et al., 2016). This acyclic nature provides propagation probabilities to an endpoint or outcome, and a BbN structure can be defined using a conceptual or influence "box and arrow" diagram (Pearl & Russel, 2001). It is only when the network includes a set of probabilities, one for each node, specifying the belief that a node will be in a particular state given the states of those nodes that affect it directly to its parents, that it becomes a complete Bayesian-belief network (Neopolitan, 2003; Norrington, Quigley, Russell, & Van der Meer, 2008).

There are several advantages of structured BbN (Chakraborty et al., 2016; Fenton & Neil, 2012; Neopolitan, 2003). First, structured BbNs minimise specifying probabilities by having influential and fewer nodes, fewer arcs, fewer states. Second, reduce expert elicitation, including potential bias, go beyond the expert knowledge base, and overrepresent poor knowledge. Third, too many detailed steps and rules in developing a BbN can decrease model accuracy.

In BbN, the states of a variable can conceivably describe any state possible in the 'real world' (Cárdenas et al., 2012). However, they must be defined as finite in number, discrete, and mutually exclusive (Pearl & Russel, 2001). States of a variable can be Boolean (e.g., true or false), categorical (e.g., high, average, low), discrete (e.g., integers) or continuous (Pearl & Russel, 2001). If a variable is continuous, then it is generally handled by dividing its range into sub-ranges with discrete values.

Therefore, the discretisation of variables is a requirement of BbNs (Chakraborty et al., 2016).

To obtain a robust and representative BbN, setting discrete intervals in a BbN should not be an arbitrary process. For this, breakpoints in data distributions should be explored. For example, plotting data distributions, undertaking multivariate statistics or classification analyses of datasets, using percentiles of data, is recommended for empirical datasets (Neopolitan, 2003; Salini & Kennet, 2007). When information is subjective, expert judgment can be used; otherwise, states represent critical regulatory thresholds if a model has a decision-making context (Pearl & Russel, 2001). For example, in our study, assessing the representativeness of the data sensitive state can be defined as “very high”, “high”, “moderate”, “low”, and “very low”. These representative states should be reviewed as part of the model evaluation process.

In defining states, BbN design's accuracy will depend on how many nodes and arcs are used to model processes and the number of discrete intervals used within each variable (Bayraktarli & Ulfkjaer, 2005; Bøttcher & Dethlefsen, 2003; Herland, Hämmäinen, & Kekolahti, 2016). The model can result in information loss by selecting too few states, whereas too many states can overcomplicate the model (Chakraborty et al., 2016). While the potential loss of information can be a benefit of the process of discretization, this loss of information is less crucial where states are used to represent management objectives or outcomes (Cárdenas et al., 2012; Heckerman, 2008).

5.1.3 Conditional probability tables

A Conditional Probability Table (CPT) describes the relationship between a child node and all its parents (Heckerman, 2008; Murphy, 1998; Neopolitan, 2003). The CPT represents the probability of being within a state, given a combination of parent states' values. Therefore, each variable's CPT size is the product of the number of states of the child node and all its parent nodes. (Crandell, Voils, & Sandelwoski, 2012). When a node has no parents, it is a root node, and it can be described probabilistically by a marginal probability distribution (Neopolitan, 2003).

The following short example in Figure shows inputting data into the CPTs for a simple Bayesian network consisting of only three nodes. In the network, nodes A and B (parent nodes) represent node C's causal factors (child node).

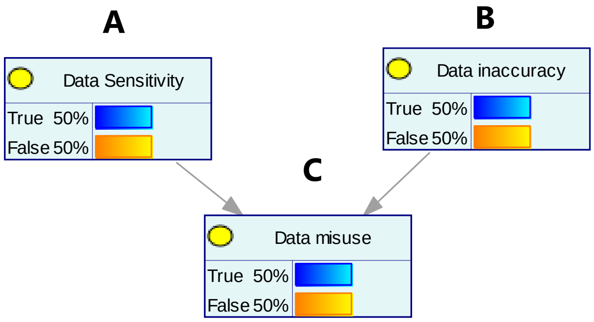


Figure 5-2 Simple model structure illustrating nodes with two states

Figure 5-2 presents that all nodes are binomial, with the states being defined as either true or false. A (Data sensitivity) variable can be described by a finite number of states, which can be defined either qualitatively or quantitatively. The probability distributions for each node have not yet been specified. Therefore, this diagram is not yet a full BN but merely a Bayesian diagram. Nodes A (data sensitivity) and B (data inaccuracy) are both root nodes. Accordingly, they can be defined by marginal probabilities. Node C (data misuse), however, is the child of A and B, and so the probabilities of the states of node C are conditional on how the states of A and B combine.

5.1.4 Methods for the BbN's propagation

Several methods are commonly used to calculate the nodes' conditional probabilities within a BbN, where possibilities can be obtained through expert elicitation (Bøttcher & Dethlefsen, 2003; Chakraborty et al., 2016). The accuracy of information obtained through elicitation can range from a deep understanding of the relationships' strength to a more heuristic estimate (Pearl & Russel, 2001). This information can also be derived from a diverse range of personal experiences of non-expert stakeholders in the system, such as anecdotal or contextual information (Murphy, 1998; Spiegelhalter, 1998; Stutz & Cheesman, 1994).

Probabilities can also be obtained by constructing equations, including probabilistic distributions, derived from fully peer-reviewed or even simple conceptual (Nadkarni & Shenoy, 2001). Furthermore, they can also be obtained from scientific data sources, including the frequency of observed conditions in the monitored field or laboratory observations and scientific surveys (Robertson & Wang, 2004). Also, due to its inherent incorporation of uncertainty, incomplete datasets can be used to calculate conditional probabilities (Neopolitan, 2003). The BbNs can use a combination of methods to calculate conditional probabilities. For instance, expert probabilities can be combined with observational data to describe outcomes of extreme events not represented in the dataset (Ben-Gal, 2008; Heckerman, 2008; Neopolitan, 2003). BbNs can use historical datasets or past experiences to quantify the probabilities of the advantages and disadvantages of the opening datasets.

Nevertheless, although it can incorporate data from a wide variety of sources, it is important to keep in mind the different types' risks and limitations (L. D. Hudson, Ware, Mahoney, & Laskey, 2002). If information is obtained from scientific data or theory, it may be incomplete or unavailable in part. Even if the information is acquired from the elicitation of professional judgement or personal experience, on the other hand, high uncertainties can arise from epistemic uncertainty, which means that the expert's judgment can be incomplete knowledge or bias (Bayraktarli & Ulfkjaer, 2005; Robertson & Wang, 2004). Therefore, as previously mentioned, all sources of information used in the creation of any model must be documented clearly and transparently (Nadkarni & Shenoy, 2001; Neopolitan, 2003; Pearl & Russel, 2001).

Chance (node)		% Probability	
A	B	True	False
True	True	100	0
True	False	75	25
False	True	50	50
False	False	0	100

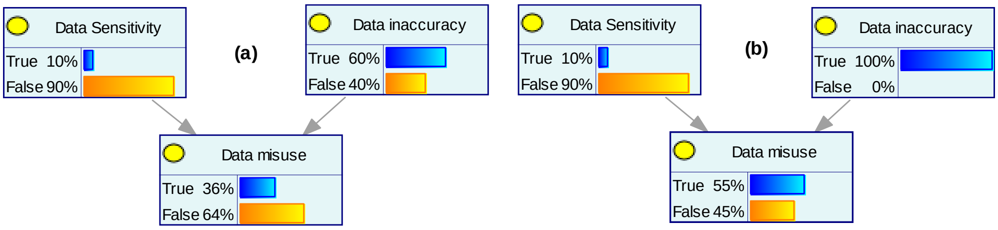


Figure 5-3 BbN before (a) and after (b) the propagation of new information

In the illustration presented in Figure 5-3, the elicitation process would usually take the form of scenarios as they appear in the table. For example, given that A (data sensitivity) is true and B (data inaccuracy) is true, what is the probability that C (data misuse) is true (represented here as 100%). The fully parameterised CPT is shown in Figure 5-3. It is important to note that the probability generation method should always be rigorously documented during the experts' elicitation process, including its assumptions and limitations.

Figure 5-3 shows the propagation probabilities of new information in the BbNs. When the probability distributions of each node have been defined, the network is able to be 'solved', as shown in Figure 7(a). After evaluation tests, the BnN is complete and can be used for scenario analysis. Thereafter, individual scenarios, such as a set of decision-makers or policy-makers (in the case of open data) interventions or observations of the system, can easily be examined.

Consequently, BbNs provide a simple way of testing a scenario, allowing the user to input evidence into a node by defining a fixed distribution at a node. The effect of the scenario can then be examined by its effect (new information) on other nodes by propagating probabilities, as illustrated in Figure 5-3. The fast propagation of information through the network is one of the main advantages of BbNs method, in that they can be used to quickly view how decisions and observed conditions at one node will affect the system entirely (Nadkarni & Shenoy, 2001; Neopolitan, 2003; Pearl & Russel, 2001; Spiegelhalter, 1998).

5.1.5 Steps to construct the BbN

To better understand the relationship and influence factors of the disadvantage in opening data, in this section, we use a BbN explanatory model as an approach by employing four main steps (Chakraborty et al., 2016). At the start, we require defining the disadvantage variables and their relationships. Second, we construct a network structure of the disadvantages to show how the variables are interrelated. In the third step, we interrogate the model to get better comprehend the vulnerability of disadvantage variables. As a final point, we develop the relationship diagrams to communicate the outcomes to the related stakeholders. There are several ways to show how to develop BbN. In this illustration, we adopted six main sequential phases to construct causality and relationships between factors influencing the advantages and disadvantages of opening data (Chakraborty et al., 2016), as visualised in Figure 5-4.

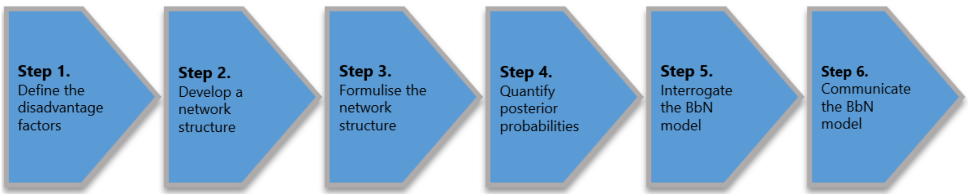


Figure 5-4 Step-by-step of the BbN development model

First, we define the disadvantages variables. Then, we classified the disadvantages into several categories to make a singular variable of the cost and disadvantage for organising the cause-and-effect elements. To make a distinct understanding and avoid misinterpretation of these categories, we described the disadvantage factors of opening data, as presented in Chapter 3.

Second, we develop a network structure to present disadvantages that influence the potential of disclosing a selected dataset. Three sub-steps should be followed to identify sub-nodes and their relationship: (1) Identify the key elements. The key elements will turn out to be parent nodes of the top-level node. This sub-step aims to develop further sub-nodes of the disadvantage variables that influence them; (2) Identify the remaining elements to describe the various risk elements' causality until the lowest level is generated; and identify the relationship between the disadvantage factors to identify the various nodes includes key elements and other related elements, based on their influence diagrams. The relationship knowledge in this work will be identified from the literature-based.

Third, we formulise the network structure. The Bayesian-belief Network is able to formulise the uncertainty in the dependencies between the defined variables using conditional probabilities (Cárdenas et al., 2012). The probability factors in the Bayesian-belief Network are also able to compute the effect of any variable from the probability of a given cause element.

Fourth, we quantify the posterior probabilities. The objective of calculating posterior probabilities is to define and estimate the probability distributions for each benefit and cost factor. There are two main procedures to quantify the posterior probability factors. First, select the experts' team based on their formal education, functional knowledge, and practical insight. Second, quantify the cost and benefit factors by the experts' judgment.

Fifth, we interrogated the model. This step is aimed to interrogate the sensitivity and influence of variables on the disadvantages. Sensitivity means that each node's responsiveness or variable in the network structure is analysed using a systematic approach to express the trigger variables (Chakraborty et al., 2016; Herland et al., 2016). Simultaneously, the influence factors tend to analyse the parent nodes' frequency of impacts on their respective child nodes by identifying their influential elements. This step's expected result is better to understand the most substantial disadvantages in opening data.

Sixth, we communicated the model. The final process of BbN development in this study is how to communicate the resulting network model. In this illustration, we utilise the relationships diagram to describe the results visually. In this step, we illustrate the relationships of influence disadvantages variable in opening data. The four steps explained in the research approach are followed. We illustrated using a medical record dataset to analyse the potential disadvantages containing inside. Supposed that medical records dataset consists of some fields such as *name_of_patient, date_of_birth, address, and phone_number.* The government would like to analyse the potential disadvantages of the fields before it released. By constructing a BbN, the government can better understand the causality and relationships factors influencing the disadvantages of opening data.

5.1.5.1 Define the disadvantage factors

Open data has been shown to contribute to society through several programs and agendas of many countries' governments in recent years (Zuiderwijk, Janssen, et al., 2014). At the same time, along with the benefits of implementing the disclosure of data, potential disadvantages of disclosing data are highlighted (Barnickel et al., 2012; M. Janssen et al., 2012; Martin et al., 2013). The disadvantages are classified into five categories: data inaccuracy, data misuse, data sensitivity, data incompleteness, and data misinterpretation. These disadvantage categories are derived from Table 3-4 in Chapter 3.

5.1.5.2 Develop a network structure

In this step, we developed a BbN structure to identify the causes and relationships between disadvantaged elements. In step 1 (define the disadvantage variables), we defined the five main disadvantage categories of opening data. Based on the cause-and-effect for each category, a BbN structure was generated. In doing so, there are three sub-steps to complete this step. (a) Identify the key elements. We classified the parent nodes from the identified disadvantage elements and then (b) identified the remaining elements. After the parent and the child nodes have been identified, we connected the parent and child as the one-to-one or one-to-many connectivities, and

(c) identified the relationships. This process's main objective is to make a relationship between parent nodes. This step needed several iterations until the lowest sub-node is identified and correlated.

A. **Identify the key elements and relationships**. We identified the parent nodes of the top-level nodes for each of the five categories.
   - Data inaccuracy. Factors such as data entry mistakes, flawed data entry process, the null problem, and deliberate error as the influencing factors are sub-elements (D. Chen & Zhao, 2012; Dekkers et al., 2014; Kucera & Chlapek, 2014).
   - Data misuse. This category's cause elements are discredited personal profile, unauthorised user, and diminish reputation (Amit & Larson, 1990; Kucera & Chlapek, 2014).
   - Data sensitivity. Releasing data can include sensitive attributes. The users can analyse personal identity elements, like full name, date of birth, address, and phone number. Therefore, data sensitivity influencing data privacy and data violation (Barry & Bannister, 2014; Kulk & van Loenen, 2012).
   - Low data quality. Opening incomplete data create misunderstanding about the quality of the data. The caused elements of this category are (a) the anonymity of the data source, (b) inappropriate aliases formula, and (c) mismatch of the attribute relationships. This situation can also influence data quality and misinterpretation (Amit & Larson, 1990; Kucera & Chlapek, 2014; Walter, 2001).
   - Data misinterpretation. Publishing data can be interpreted in the wrong ways. Possible cause factors of this category are: (a) insufficient domain expertise, (b) essential variables are omitted, (c) inappropriate data visualisation, and (d) error of attribute correlation. This disadvantage category's effect influences the data quality and data incompleteness (Barnickel et al., 2012; Barry & Bannister, 2014).

B. **Identify the remaining elements**. From the parent node and sub-elements constructed, we generated the connection between the variables. The correlation of each node and sub-elements shows the relationship between the disadvantage elements until the lowest level, as presented in Figure 5.5.

C. **Identify relationships**. In this step, the consequence elements of the parent nodes are constructed. In Figure 5-5, the developed Bayesian network relationship is shown. The parent nodes (open dataset) influence three main disadvantage factors, namely data sensitivity, data misinterpretation, and low data quality.

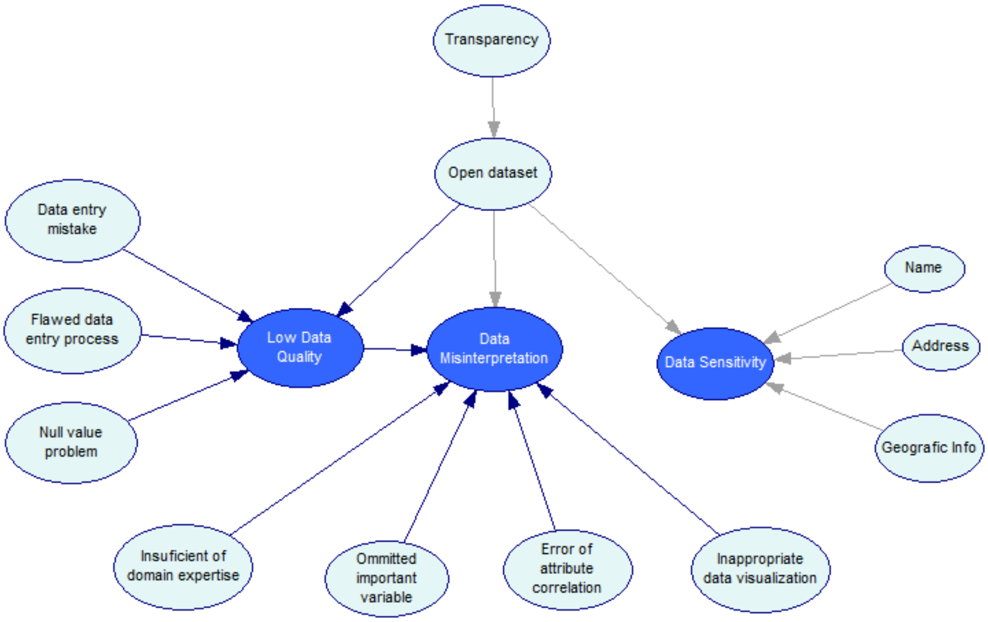


Figure 5-5 Identify key elements and relationships

5.1.5.3 Formulise the network structure

The expert's judgment in this formulation refers to experts' subjective prior beliefs about the potential costs and benefits of opening data. The formulation to compute the probabilities of the cost and benefits factors of opening data is defined as follows:

P[effect]=[P[effect/cause].P[cause]]/P[cause/effect]

Where:

P[cause] = probability that the cause occurs,
P[effect] = probability that the effect occurs,
P[effect/cause] = conditional probability of the effect, given the cause,
P[cause/effect] = conditional probability of the cause, given the effect.

To illustrate how to formalise the disadvantage factors in opening data, it can be shown as follows:

Potential disadvantage	= [P(D)] = Evidence;		Expert_belief	= [Bp];
Data sensitivity	= [P(T)];		Interview	= [Vp].
Data misinterpreation	= [P(A)];			
Low data quality	= [P(L)];			

(1)	P(D|T) P (data sensitivity)	$= \sum_{i=1} P(T \mid Bp_i)\ P\ (Vp_i)$ $= \sum_{i=1} P(Data\_sensitivity \mid Expert\_belief_i)\ P\ (Interview_i)$
(2)	P(D|A) P(data misinterpretation)	$= \sum_{i=1} P(A \mid Bp_i)\ P\ (Vp_i)$ $= \sum_{i=1} P(Data\_inaccuracy \mid Expert\_belief_i)\ P\ (Interview_i)$
(3)	P(D|L) P(low data quality)	$= \sum_{i=1} P(L \mid Bp_i)\ P\ (Vp_i)$ $= \sum_{i=1} P(low\_data\_quality \mid Expert\_belief_i)\ P\ (Interview_i)$
(4)	P(D) P(disadvantage)	$= \sum P(T, A, L \mid Vp_i) \times P\ (Vp_i)$

5.1.5.4 Quantify posterior probabilities

In this step, we quantify the posterior probabilities for each disadvantage factor to estimate the probability distributions. There are two steps to quantify the posterior probability factors: (a) select the experts based on their formal education, functional knowledge, and practical insight; (b) quantify the cost and benefit factors by the experts' judgment.

a. Experts' domain and expertise

Ideally, the experts need to accommodate various specializations that partially overlap to confirm the completeness of the data or information available (Herland et al., 2016; Teicher, 2015). the experts' selection is based on their formal education, functional knowledge, and practical insight. In this study, there were three experts involved from various domains contributing to the probability quantification process. The experts come from academia (a doctoral student in the open government data research interest), from the government (a computer security analyst to provide expertise in the field of privacy knowledge and issues), and from the community (a public dataset analyst to use expertise and view related to transparency for opening dataset to the public domain). Ideally, we need more experts in helping the quantification process. However, there was not possible to find experts who had a strong knowledge of the open government data domain and could estimate the potential disadvantages of opening data.

The experts need to accommodate various specializations that partially overlap to confirm the completeness of the data or information available (Herland et al., 2016; Teicher, 2015). Finally, we interviewed experts from the best practice insights. The interviewee's expertise and experiences in the knowledge domain must be high to warrant the quality and validity of available information (Herland et al., 2016; Honda, Washida, Sudo, Wajima, & Awata, 2017).

b. Experts' Judgment quantification

Expert judgment quantification results in numerical data form representing the event frequencies, causal relationship, and conditional probabilities in terms of disadvantages for opening data. We highlight two requirements that regulate how the selected experts (Little & Cooke, 2016; Oslon, 2010). First, the experts quantify the approximate's relative deviation from the mean of all experts' estimates of the selected case study. There was not always consensus among the experts. For example, Table 5-1 shows different views from the experts regarding the level of disadvantages. Expert 1 and 2 believed that releasing a geographic information attribute will probably affect a moderate disadvantage level (0.59 and 0.54). In contrast, expert 3 was convinced that this attribute had only a low level of disadvantage (0.42). There was no consensus among experts in this case, and we took the mean score from the three experts. Second, we expect the selected experts to consistently quantify the cost and benefit factors to offer the high, moderate, and low-status factors.

Table 5-1 Quantification of the disadvantage factors

Expert 1.

Factor	Sub-factor		Probability Quantification		
			High	Moderate	Low
Data sensitivity (T)	T.1	Name	0.45	0.32	0.23
	T.2	Address	0.35	0.55	0.10
	T.3	Geographic info	0.28	0.59	0.13
Data misinterpretation (A)	A.1	Insufficient of domain expertise	0.28	0.64	0.08
	A.2	Omitted important attribute	0.24	0.54	0.22
	A.3	Error of attribute correlation	0.24	0.64	0.12
	A.4	Inappropriate data visualisation	0.36	0.54	0.10
Low data quality (L)	L.1	Data entry mistake	0.34	0.30	0.36
	L.2	Flawed data entry	0.25	0.44	0.31
	L.3	Null value problem	0.35	0.44	0.21

Expert 2.

Factor	Sub-factor		Probability Quantification		
			High	Moderate	Low
Data sensitivity (T)	T.1	Name	0.42	0.28	0.30
	T.2	Address	0.22	0.44	0.34
	T.3	Geographic info	0.26	0.54	0.20
Data misinterpretation (A)	A.1	Insufficient of domain expertise	0.32	0.41	0.27
	A.2	Omitted important attribute	0.18	0.48	0.34
	A.3	Error of attribute correlation	0.24	0.55	0.21
	A.4	Inappropriate data visualisation	0.24	0.44	0.32
Low data quality (L)	L.1	Data entry mistake	0.23	0.25	0.52
	L.2	Flawed data entry	0.27	0.39	0.34
	L.3	Null value problem	0.42	0.47	0.11

Expert 3.

Factor	Sub-factor		Probability Quantification		
			High	Moderate	Low
Data sensitivity (T)	T.1	Name	0.55	0.14	0.31
	T.2	Address	0.36	0.42	0.22
	T.3	Geographic info	0.33	0.25	0.42
Data misinterpretation (A)	A.1	Insufficient of domain expertise	0.54	0.31	0.15
	A.2	Omitted important attribute	0.23	0.44	0.33
	A.3	Error of attribute correlation	0.47	0.47	0.06
	A.4	Inappropriate data visualisation	0.29	0.44	0.27
Low data quality (L)	L.1	Data entry mistake	0.27	0.25	0.48
	L.2	Flawed data entry	0.47	0.39	0.14
	L.3	Null value problem	0.49	0.47	0.04

Mean score from Expert 1, 2, and 3

Factor	Sub-factor		Probability Quantification (Mean score)		
			High	Moderate	Low
Data sensitivity (T)	T.1	Name	0.47	0.25	0.28
	T.2	Address	0.31	0.47	0.22
	T.3	Geographic info	0.29	0.46	0.25
Data misinterpretation (A)	A.1	Insufficient of domain expertise	0.38	0.45	0.17
	A.2	Omitted important attribute	0.22	0.49	0.29
	A.3	Error of attribute correlation	0.32	0.55	0.13
	A.4	Inappropriate data visualisation	0.30	0.47	0.23
Low data quality (L)	L.1	Data entry mistake	0.28	0.27	0.45
	L.2	Flawed data entry	0.33	0.41	0.26
	L.3	Null value problem	0.42	0.46	0.12

5.1.5.5 Interrogate the BbN model

This step aims to interrogate the disadvantage elements' sensitivity level and present the high, moderate, and low disadvantages. After constructing the BbN causality, we interrogated the resulting model by distributing each node's probabilities and sub-elements. To integrate the sensitivity level's value, we used the experts to quantify each disadvantage element's probability. We expressed the value of the possibilities into three-level, namely "High", "Moderate", and "Low". We used three levels as this would keep it simple and easy to use by the experts. The high level refers to the node will likely have multiple severe adverse effects for opening data. The moderate level means that the dataset will likely have a moderate adverse effect, and the low level of the disadvantage refers to the dataset will likely have a limited adverse impact on the disclosing data. Since the experts' levels of knowledge and expertise were diverse, providing a simple assessment matrix shown in Table 5.1 could help the expert against time issues.

The objective of the expert's quantification is to collect the information necessary to construct the quantitative Bayesian network model, which is the set of Node Probability Tables (NPT) in Table 5-1 assigned to the nodes of the nodes qualitative Bayesian network. Figure 5-6 presents the causality and relationships between factors influencing the advantages and disadvantages of opening data. This figure shows the complete BbN structure, including the nodes' state probability distributions. The experts quantified both cause and effect nodes. For example, the disadvantages of opening a dataset can directly cause three main possible risks, namely data sensitivity, data misinterpretation, and low data quality. Figure 5-6 also visualises the probability of occurrence for each disadvantage, which shows that most risks are unlikely to occur during data opening. Based on Figure 5-6, experts believe that 50% of the opening dataset becomes high risks of data sensitivity, followed by 46% of data misinterpretation and 44% of low data quality, respectively. Furthermore, data sensitivity is influenced by three sub-nodes of the dataset attribute: name, address, and geographic info. In Figure 5-6, experts believe that name can result in high risk (47%), while address and geographic info can moderately influence the data sensitivity with percentages of 47% and 46%, respectively.

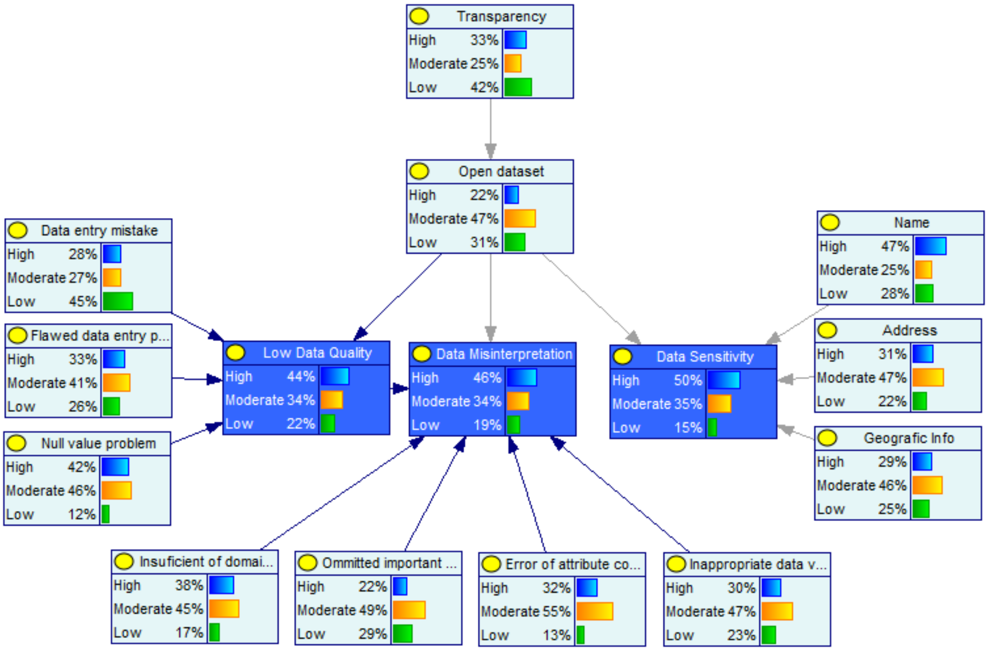


Figure 5-6 Interrogating the disadvantage factors

5.1.5.6 Communicate the model

The final step of the BbN model is to create a model that can be communicated to stakeholders like decision-makers, policy-makers, and data providers. There are some approaches to disseminate the data to the public, like graphs, charts, histograms, or scatter plots can help the stakeholders use the models in practice (Chakraborty et al., 2016). The communication of the model for the quasi-experiment will be explained in detail in Chapter 7.

To conclude, the public expects government organisations and data providers to open their data for gaining benefits. However, governments are often risk-averse due to possible disadvantages. The causality and relationships between factors influencing the advantages and disadvantages of opening data are not investigated yet in the literature. Therefore, we used a BbN to construct a causal model of the disadvantages in opening data.

5.2 Fuzzy Multi-criteria Decision Making

Fuzzy Multi-Criteria Decision Making (FMCDM) is a method to determine the best alternative of a decision problem and manage the decision-making problem of alternative selection (Hsieh, Lu, & Tzeng, 2004). These alternatives are developed by establishing and incorporating the FMCDM based on the Fuzzy Analytic Hierarchy Process (FAHP) (Hsieh et al., 2004; Rezaei et al., 2013). The Fuzzy logic's primary function is to capture the expertise of open experts and express it with a computational approach (Fuller, 1999; Gupta, 1995; Zadeh, 1975).

Furthermore, Fuzzy theory is based on intuitive reasoning by considering human subjectivity and incorrectness, which are common in the natural language (Werro, 2015). The natural language is an intricate structure both in human communication and how the human being thinks (Novák, 1992; Werro, 2015). A fuzzy theory provides a numerical strength for the emulation of the higher cognitive function from the human thought and perception associated with weights of the advantages and disadvantages of opening data.

The important role of the FMCDM is to assess the alternative selection related to predetermined criteria for a single decision making (Kahraman et al., 2015). The appropriateness of the alternative compares to the requirements, and the priority weights of each measure can be analysed and computed using linguistic matrix values reflected by the fuzziness (S.-J. Chen & Hwang, 1992; Zadeh, 1975). FAHP was utilised to determine the preference weightings of criteria by collecting expert's judgment (Hancerliogullari, Oymen, & Koksalmis, 2017; Hsieh et al., 2004). The scores for each criterion are summed up to rank the importance of the alternatives (Lin & Twu, 2012; Sloane, Liberatore, & Nydick, 2011).

This FAHP technique consists of six following steps (Hancerliogullari et al., 2017; Hsieh et al., 2004; Rezaei et al., 2013). First, select experts to help in scaling the fuzzy linguistics matrix. Second, determine the evaluation criteria and construct the hierarchy, including their alternatives. Third, construct a pairwise comparison matrix and evaluate the relative importance of the criteria. Fourth, transform the linguistic terms into the triangular fuzzy number. Fifth, calculate the Fuzzy weights matrix and check the pairwise comparison matrix's consistency. Sixth, select the best alternative of a decision. A dataset of medical records is used in the illustration part to show how the advantages and disadvantages of multiple criteria can be analysed by employing the FMCDM approach.

Furthermore, we proposed four possible decisions regarding the decision-making process's final steps: open, maintain data suppression, provide limited access, and remain closed the dataset. These alternative decisions will be analysed based on the four main disadvantage criteria: data sensitivity and data ownership. At the same time, we also provide data availability and data trustworthiness to the benefit criteria.

Data sensitivity and ownership are selected as input because these criteria can represent privacy violation issues containing the medical record dataset. For example, in the case of data sensitivity, by releasing the actual name, date of birth, place of birth, home address, or the insurance provider of a patient, these might be potentially misused by unauthorised users. Simultaneously, data availability and data trustworthiness are chosen criteria because they reflect the advantages of transparency and accountability in opening data. Both of the requirements have sub-criteria to refine the advantages and disadvantages further.

5.2.1 Steps of the FMCDM

To explain how the FMCDM works to support decision-making, we use three main detailed decision-making sub-steps: data source, evaluation, and decision. The whole process starts with selecting the data source dataset to create the input for the evaluation phase. The input data are processed next in the evaluation phase. The output of the evaluation, namely the alternative decision stage, results in a suggestion to make a decision. The latter is done by showing the rank of decision priority, as shown in Figure 5-7.

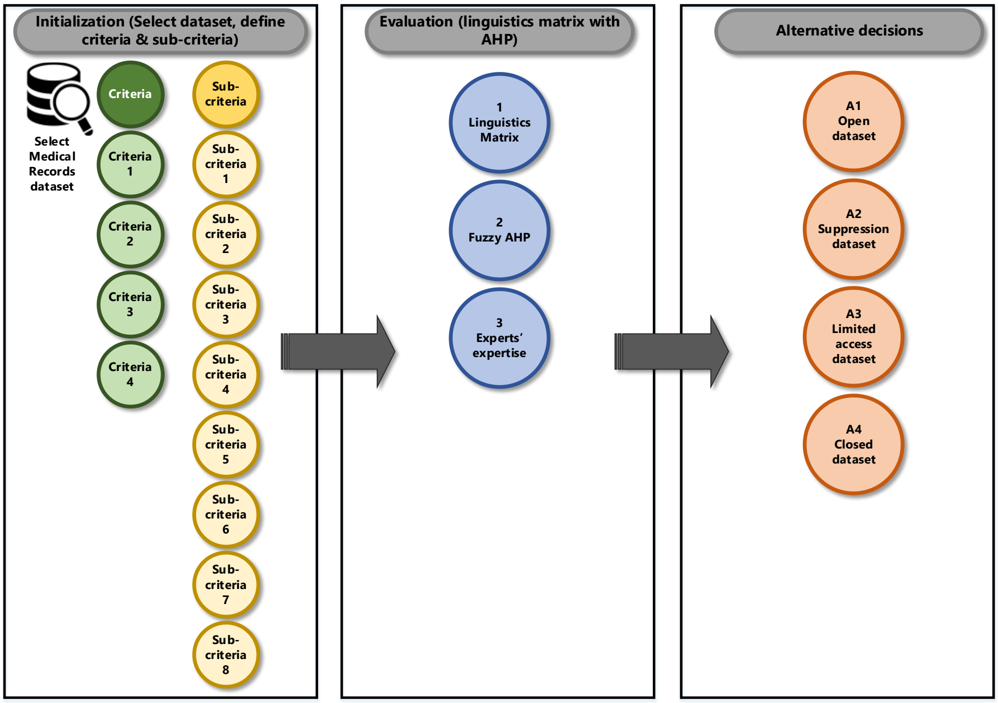


Figure 5-7 The flow process of FMCDM

The decision-making process's steps to open data using the FMDM method is based on three main sub-steps. First, we initialise the process of selecting and extracting the medical records dataset. Next, we defined the four criteria and another eight sub-criteria. Second, we evaluated the criteria dan sub-criteria of the potential advantages and disadvantages of opening data by using three steps: developing a linguistics matrix, employing fuzzy AHP, and quantifying the fuzzy linguistics scale using an expert's judgment. Third, we provide alternative decisions about whether to open, introduce the dataset's suppression, provide limited access, and remain closed the data. Figure 5-7 illustrates the flow process of FMCDM to analyse the advantages and disadvantages of opening data.

A. Initialisation. First, we need to select the type of dataset. For example, in this case, we retrieved the medical records dataset. We use the diagnosed stage table as the primary table to analyse (see Table 5-2). Next, we defined the criteria and sub-criteria of the advantages and disadvantage factors using the taxonomy. The taxonomy of advantages and disadvantages to open data derived from the systematic literature study we carried out, as provided in the previous chapter (see Chapter 3, Figure 3-2). In our illustration, four criteria and eight sub-criteria of the advantages and disadvantages serve as input data.

B. **Evaluation.** In the second step, we used FMCDM to evaluate the alternatives based on criteria defined in the data source elicitation phase. The criteria use linguistic matrix values as presented in Table 5-3 reflected by the Fuzzy Linguistic Scale. FMCDM works on the Fuzzy AHP technique, which has an essential role in measuring the relative importance of defined criteria for decision-making problems. Moreover, to quantify the relative importance of the advantages and disadvantages, we selected the knowledge from experts' judgment. There are two main steps to conduct an evaluation process by the experts in AHP (Hancerliogullari et al., 2017; Podvezko, 2011). First, experts should rank the criteria in a descending or ascending order of their significance. Then, determining the most important criteria and compare them with others. For example, an expert ranked that data sensitivity (C1) is higher or essentially important than data ownership (C2). Second, experts will determine the weights by transforming a pairwise comparison matrix into a triangular fuzzy number.

C. **Decision**: Finally, this flow process's outcome is to get the best alternative's final weights as the priority of a decision.

5.2.2 Decision alternatives

There are four decision alternatives of opening data provided in this illustration, namely opening the dataset (A1), maintaining a dataset suppression (A2), providing limited access (A3), and remaining closed the dataset (A4). Opening the dataset refers to publishing the dataset that might present a small disadvantage to an individual or organisation identity. At the same time, the potential benefits of the dataset substantially outweigh the potential disadvantages. Second, maintaining suppression to the dataset refers to removing a data field and individual records into particular groups or generate unique characteristics to avoid re-identification. In this alternative, data that might create significant disadvantages are not opened as the potential benefits do not outweigh the possibility of the disadvantages. Limited access to the dataset defines that only a particular group of users can access the data. In this decision, the level of openness is limited, and often, those who will gain access have to sign a document that outlines the rules of entry. This is because releasing the dataset will create a moderate disadvantage, or the potential benefits of the dataset do not outweigh the potential privacy disadvantages. The remaining closed decision to the dataset means that publishing the dataset generates a very high disadvantage to an individual or organisation and significantly outweighs the potential benefits.

5.2.3 Selection criteria

Figure 5-8 illustrates the hierarchy of the four criteria, eight sub-criteria, and four alternatives. The four criteria (C1, C2, C3, and C4) define data sensitivity, ownership, availability, and trustworthiness. The data sensitivity (C1) composes of two sub-criteria, namely individual life-threatening (C1.1) and data identifiable (C1.2). Individual life-threatening (C1.1) can be defined as a potential disadvantage to an individual or personal life because of the possibility of recognising the dataset's sensitive value. Data identifiable (C1.2) is specified as the possible leak of the personal, organisational, business, or even organisational data identity, e.g., by combining some table attributes.

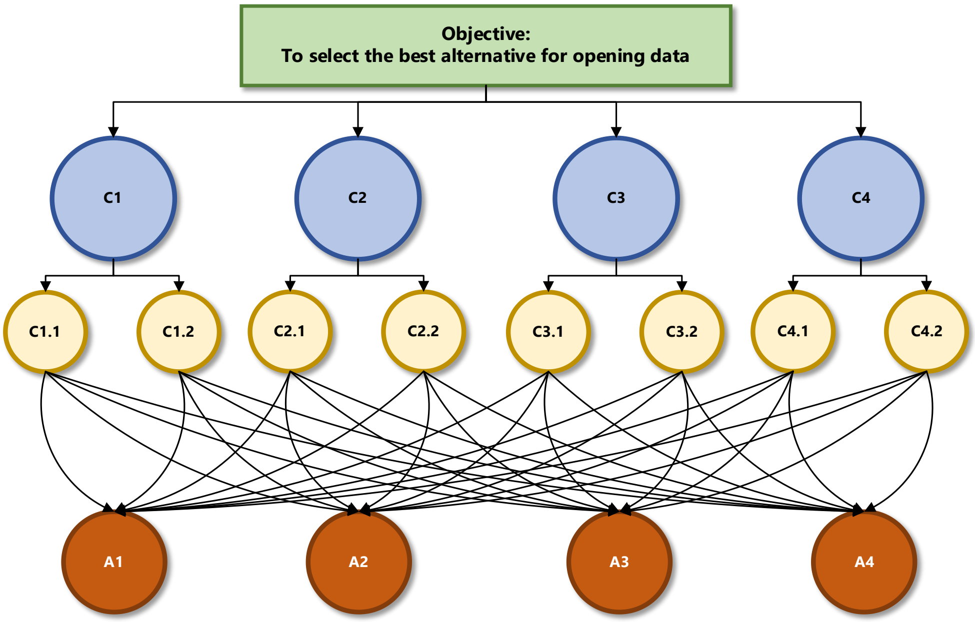


Figure 5-8 Illustration of the hierarchy of criteria and alternatives

The second criterion is data ownership (C2), which consists of two sub-criteria, namely metadata scanning (C2.1) and fake or misleading (C2.2). Metadata scanning (C2.1) can be used to figure out the property and structure of the dataset. Fake or misleading (C2.2) is defined by a user to potentially change and modify the dataset and affect an unreliable and wrong decision. Data availability (C3) is the third criterion, and it has two sub-criteria, namely, data manageability (C3.1) and data recoverability (C3.2). Data manageability (C3.1) is specified as the chance to manage the dataset's availability and accessibility. Data recoverability (C3.2) is the ability to re-construct a published dataset. The fourth criterion is data trustworthiness (C4), which consists of two sub-criteria, e.g., data traceability (C4.1) and data authenticity (C4.2). Data traceability (C4.1) can make the possibility to trace the source of the dataset. Data authenticity (C4.2) is defined as potentially affected to recognise the authentication of the data.

5.2.4 Fuzzy AHP technique

The AHP process is a quantitative method that hierarchically deals with multi-attribute, multicriteria, and multi-period problems (Saaty, 1980). Only with AHP, it is impossible to overcome the fuzziness's deficiency during decision-making (Kuo, Liang, & Huang, 2006). Hence, in this study, the Fuzzy AHP is the extension of the conventional AHP method by integrating fuzzy comparison ratios for multicriteria analysis (Hancerliogullari et al., 2017; Hsieh et al., 2004; Isselhardt & Cappuci, 1989; Saaty, 1980). It uses the triangular fuzzy number of fuzzy set theory directly into the pairwise comparison matrix of the AHP. The geometric mean method is used to generate fuzzy weights and performance scores (Sehra, Brar, & Kaur, 2012). The steps of the Fuzzy AHP can be summarised as follows:

- Step 1. Select experts. The evaluation process is abed on the experts' knowledge and experience. Hence the selection of experts is crucial.
- Step 2. Determine the evaluation criteria and construct the hierarchy, including alternatives.
- Step 3: Construct a pairwise comparison matrix and evaluate the relative importance of the criteria. The experts are expected to provide their judgment on the basis of their knowledge.

  For any expert, the comparison matrix is given by Eq. (1) as:

a) $\tilde{C}_k = \begin{bmatrix} 1 & \tilde{c}12 \cdots & \tilde{c}1n \\ \vdots & \ddots & \vdots \\ \tilde{c}n1 & \tilde{c}n1 \cdots & 1 \end{bmatrix}$

(1)

where n is the number of criteria, $\tilde{C}_k$ is a pairwise comparison matrix belongs to k^{th} expert for k=1, 2. k.

The arithmetic mean is used to aggregate experts' opinions as given in eq.

(2).

b) $\tilde{C} = \frac{1}{k}\left(\frac{1}{C} + \frac{2}{C} + \cdots + \frac{k}{C}\right)$

- Step 4: Transform the linguistic terms into fuzzy triangular numbers. The following linguistic terms provided in Table 2 are utilised for the evaluation procedure.
- Step 5: Calculate the fuzzy weight matrix using eq. (3) and eq. (4).

  $\tilde{r}_i = (\tilde{c}_{i1} \otimes \tilde{c}_{i2} \otimes \ldots \otimes \tilde{c}_{in})^{\frac{1}{n}}$

  (3)

  $\widetilde{w}_i = \tilde{r}_i \otimes (\tilde{r}_1 + \tilde{r}_2 + \cdots + \tilde{r}_n)^{-1}$

  (4)

  where $\tilde{r}_i$ is the geometric mean of fuzzy comparison value and $\widetilde{w}_i$ is the fuzzy weight of the i^{th} criteria.

- Step 6: Apply the normalisation procedure as eq. (5)

  $w_i = \frac{\widetilde{w}_i}{\sum_{j=1}^{n} \widetilde{w}_j}$

5.2.5 Analyse dataset using FMCDM

Likewise, to the previous method (BbN), we demonstrated the FMCDM using the medical records dataset with the Fuzzy AHP technique's help. The reason for selecting this dataset is that it contains both advantages and disadvantages categories. The variety of benefits from the selected dataset includes the availability of hospital medical records by providing accurate, up-to-date data and enabling quick access to the patient records. However, by releasing the patient health records attributes, data like the *name_of_patient, date_of_birth, and place_of_birth* might be opening, resulting in a privacy violation.

Step 1. Data Source: Medical Records Dataset

For the illustration, the Department of Health wants to release a dataset of the patient's medical records to the public that can enable individuals or organisations to access and see the current trend of a disease (Bøttcher & Dethlefsen, 2003; Kostkova et al., 2016). By doing so, a location map related to the disease landscape for some regions can be created.

However, suppose the government decides to open the complete dataset. In that case, some potential privacy issues might be very harmful (Bøttcher & Dethlefsen, 2003; Ozair, Jamshed, Sharma, & Aggarwal, 2015; Spooner & Pesaturo, 2013). Table 5-2 shows the health patient records dataset structure that will be analysed using FMCDM in this study.

Table 5-2 Row tables of Medical Records dataset

Table A (Diagnosed Stage)	Table B (Undergoes Surgery)	Table C (Metastatic Disease)	Table D (eGFR)	Table E (1L Therapy)	Table F (Hospitalised)
Name of patient	Date of surgery	Metastases sites	Biopsy date	Regimen name	Date_of_hospitalised
Date of birth	Surgeon	Time to recurrence	Test of date	Duration of therapy	Cost_of_care
Place of birth	Anaesthetic		Turnaround time	Dosage	
Gender			Number of unsuccessful tests	Cocominant meds	
Race			Test result	Response	
Insurance			Type of eGFR	Line of therapy	
Stage				Laboratory name	
TNM staging					

In this work illustration, we use Table A, namely Diagnosed Stage, which contains eight attributes: *Name_of_patient, Date_of_birth, Place_of_Birth, Gender, Race, Insurance, Stage, and TNM_staging.*

Step 2. Evaluation: Analysing the Dataset

The following sub-steps are the scenarios of FMCDM. Figure 5-9 shows the hierarchy of criteria and alternatives are used in the illustration of FMCDM. There are six sub-steps in this evaluation process, namely (1) establish an expert team, (2) determine the evaluation criteria, (3) construct a pairwise comparison matrix, (4) transform the linguistic terms into fuzzy triangular numbers, (5) calculate the fuzzy weight matrix, and (6) apply normalisation procedure.

In the first step, we collect the data from experts. The experts have a varied educational background, experiences in the open government data area, and best practices in estimating the potential benefits and disadvantages of opening data.

Next, in the second step, we determine the evaluation criteria of the disadvantage factors and construct the hierarchy of the criteria to break down the disadvantage factors and identifying four decision alternatives, e.g., “open”, “suppression”, “limited access”, and “closed” the dataset.

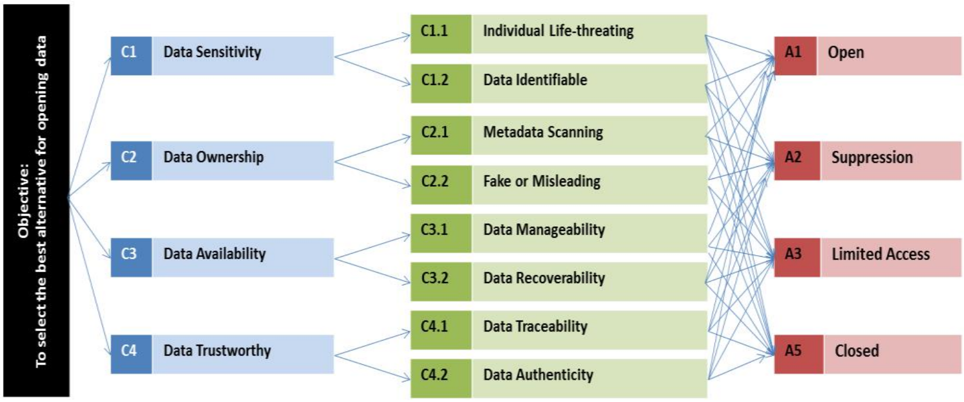


Figure 5-9 Hierarchy of criteria and alternatives for the illustration

In the third step, we constructed a pairwise comparison matrix and evaluated the relative importance of the criteria. We asked the experts to provide their consideration based on their knowledge and expertise. Only a simple pairwise comparison matrix for one expert is given for this illustration, as shown in Figure 5-9. A Fuzzy evaluation linguistic scale for the weights was constructed before the experts started quantifying the criteria, together with the experts, as presented in Table 5-3.

Table 5-3 The Fuzzy linguistic scales (adapted from: (Hsieh et al., 2004))

Fuzzy Number	Linguistic Scales	The scale of Fuzzy Number
1	Equal Important (EI)	(1,1,3)
3	Weakly Important (WI)	(1,3,5)
5	Essentially Important (SI)	(3,5,7)
7	Very Strongly Important (VI)	(5,7,9)
9	Absolutely Important (AI)	(7,9,9)

In the fourth step, we transformed the linguistic terms into fuzzy triangular numbers. The linguistic terms provided in Table 5-3 are utilised for the evaluation procedure. Next, we calculated the fuzzy weight matrix using Eq. (3) and Eq. (4) in the fifth step. The final weights of the alternatives were calculated using Eq. (3), (4), and (5). The linguistic terms provided in Table 5-2 were utilised for the evaluation, and a fuzzy qualitative approach is used for the calculation (Hancerliogullari et al., 2017; Hsieh et al., 2004). Illustrative examples for the weights of sub-criteria C11 and C12 are given as follows:

Calculating sub-criteria: Linguistic terms for the pairwise comparison from Figure 5-10 and the corresponding fuzzy numbers from Table 5-3 are used. The pairwise comparison of (C1.1 C1.2) is "Equal Important", and the fuzzy number of this linguistic term is (1,1,3).

$$\tilde{r}_{c11} = (\tilde{c}_{c11c11} \otimes \tilde{c}_{c11c12})^{\frac{1}{2}}$$
$$\tilde{r}_{c11} = \left((1,1,1) \otimes (3,5,7)\right)^{\frac{1}{2}}$$
$$\tilde{r}_{c11} = (1.73,2.23,2.64)$$
$$\tilde{r}_{c12} = (\tilde{c}_{c12c11} \otimes \tilde{c}_{c12c12})^{\frac{1}{2}}$$
$$\tilde{r}_{c12} = \left((1/(3,5,7)) \otimes (1,1,1)\right)^{\frac{1}{2}}$$

$\tilde{r}_{c12} = (0.37,0.44,0.57)$

Calculating weights: For calculating weights, we are using eq. 4. In the previous step, we are getting the value of $\tilde{r}_{c1.1}$ and $\tilde{r}_{c1.2}$ and putting these values in the following equation.

$$\tilde{w}_{c1.1} = (0.36,0.5,1.10)$$

$$\tilde{w}_{c1.2} = \tilde{r}_{c1.2} \otimes (\tilde{r}_{c1.1} + \tilde{r}_{c1.2})^{-1}$$

$$\tilde{w}_{c1.2} = (0.57,1,1) \otimes [(1,1,1.73) + (0.57,1,1)]^{-1}$$

$$\tilde{w}_{c1.2} = (0.2,0.5,0.63)$$

Expert Judgement

Criteria	C1	C2	C3	C4		C1	C11	C12	C2	C21	C22	C3	C31	C32	C4	C41	C42		
C1	1	SI	SI	SI		C11	1	EI	C21	1	SI	C31	1	EI	C41	1	EI		
C2	1/SI	1	SI	SI		C12	1/EI	1	C22	1/SI	1	C32	1/EI	1	C42	1/EI	1		
C3	1/SI	1/SI	1	EI															
C4	1/SI	1/SI	1/EI	1															
C11	A1	A2	A3	A4	C12	A1	A2	A3	A4	C21	A1	A2	A3	A4	C22	A1	A2	A3	A4
A1	1	WI	WI	WI	A1	1	WI	WI	WI	A1	1	WI	WI	WI	A1	1	WI	WI	WI
A2	1/WI	1	VI	SI	A2	1/WI	1	VI	VI	A2	1/WI	1	VI	AI	A2	1/WI	1	AI	VI
A3	1/WI	1/VI	1	SI	A3	1/WI	1/VI	1	SI	A3	1/WI	1/VI	1	SI	A3	1/WI	1/AI	1	SI
A4	1/WI	1/SI	1/SI	1	A4	1/WI	1/VI	1/SI	1	A4	1/WI	1/AI	1/SI	1	A4	1/WI	1/VI	1/SI	1
C31	A1	A2	A3	A4	C32	A1	A2	A3	A4	C41	A1	A2	A3	A4	C42	A1	A2	A3	A4
A1	1	SI	SI	SI	A1	1	EI	EI	SI	A1	1	SI	SI	SI	A1	1	EI	EI	SI
A2	1/SI	1	VI	VI	A2	1/EI	1	VI	VI	A2	1/SI	1	VI	VI	A2	1/EI	1	VI	VI
A3	1/SI	1/VI	1	SI	A3	1/EI	1/VI	1	SI	A3	1/SI	1/VI	1	SI	A3	1/EI	1/VI	1	SI
A4	1/SI	1/VI	1/SI	1	A4	1/SI	1/VI	1/SI	1	A4	1/SI	1/VI	1/SI	1	A4	1/SI	1/VI	1/SI	1

Figure 5-10 The pairwise comparison matrices of criteria and alternatives

Finally, in the sixth step, we applied the normalisation procedure, as shown below.

To find the normalised weights of C1.1 and C1.2, we used eq. 5.

$$w_{c1.1} = \frac{\tilde{w}_{c1.1}}{\sum_{j=1}^{2} \tilde{w}_{1j}} = \frac{L_{c1.1} + M_{c1.1} + U_{c1.1}}{\tilde{w}_{c1.1} + \tilde{w}_{c1.2}}$$

$$w_{c1.1} = \frac{(0.36 + 0.5 + 1.10)}{(0.36 + 0.5 + 1.10 + 0.2 + 0.5 + 0.63)} = 0.59$$

$$w_{c1.2} = \frac{\widetilde{w}_{c1.2}}{\sum_{j=1}^{2} \widetilde{w}_{1j}} = \frac{L_{c1.2} + M_{c1.2} + U_{c1.2}}{\widetilde{w}_{c1.1} + \widetilde{w}_{c1.2}}$$

$$w_{c12} = \frac{(0.2 + 0.5 + 0.63)}{(0.36 + 0.5 + 1.10 + 0.2 + 0.5 + 0.63)} = 0.40$$

A similar calculation approach is applied for all pairwise comparisons. The final weights of the alternatives are provided in Table 5-4. An illustrative example of W_{A1} is given as follows:

$$W_{A1} = C1 \times C11 \times A1 + C1 \times C12 \times A1 + \cdots + C4 \times C41 \times A1 + C4 \times C42 \times A1$$

$$W_{A1} = 0.53 \times 0.59 \times 0.39 + 0.53 \times 0.40 \times 0.41 + \cdots + 0.07 \times 0.59 \times 0.44 + 0.07 \times 0.40 \times 0.35$$

$$W_{A1} = 0.34$$

Table 5-4 Final weights of the criteria and alternatives

	C1		**C2**		**C3**		**C4**		
	0.53		0.25		0.13		0.07		
	C1.1	C1.2	C2.1	C2.2	C.31	C3.2	C4.1	C4.2	
	0.59	0.40	0.82	0.17	0.59	0.40	0.59	0.40	Weight
A1	0.39	0.41	0.41	0.41	0.44	0.35	0.44	0.35	0.34
A2	0.40	0.39	0.82	0.83	0.23	0.44	0.23	0.44	0.43
A3	0.06	0.13	0.26	0.13	0.08	0.15	0.08	0.15	0.08
A4	0.05	0.05	0.10	0.05	0.22	0.05	0.22	0.05	0.06

5.2.6 Findings of the FMCDM steps

In order to present the recommendations based on the final results of the analysing process using FMCDM, we designed a graphical view to support the decision-makers to decide to release their dataset. Figure 5-11 shows how the figure based on Fuzzy AHP can help decision-makers understand the comparison score for each alternative.

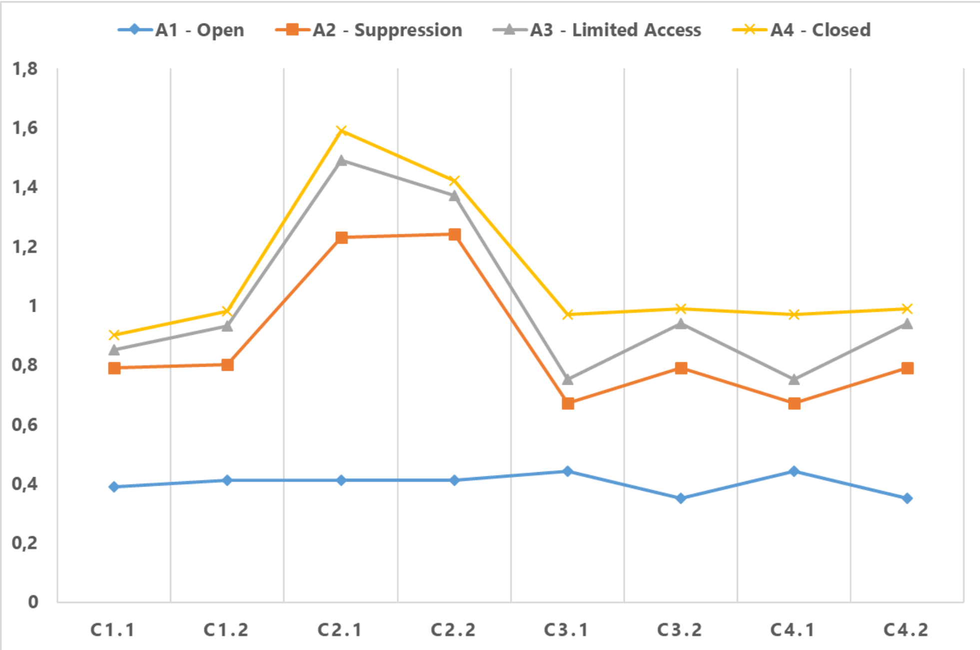


Figure 5-11 Ranking of decision recommendations

In addition, to follow up on the decision recommendation, we provided an action plan to reduce the disadvantages. Several possible alternatives are possible like (1) removing a data field or individual attributes; (2) obscuring a data field by making substitution precise data values with ranges to minimise the disadvantage of re-identification; and (3) aggregating data fields by summarising the data across the amounts of the data and visualising the data value into statistics forms, like graphics or charts.

The contribution resulted from this study is to provide a decision-making model to analyse the potential advantages and disadvantages of opening data. A given dataset is evaluated by taking action, such as measuring and weighing the multiple criteria' relative importance. Thus, the approach should support decision-makers to decide on opening datasets. In further research, we recommend refining this approach by adding more datasets in which and advice for (not) opening data can be generated without human involvement.

5.3 Decision Tree Analysis

Decision Tree Analysis (DTA) is introduced in the nineteen sixties and was primarily used in the data mining domain (Quinlan, 1990). This method's primary role is to establish classification systems based on multiple covariates in developing a prediction of alternative variables (Delgado-Gómez, C.Laria, & Ruiz-Hernández, 2019; Song & Lu, 2015). This method allows an individual or organisation to trade-off possible actions against another action based on the probabilities of advantages and disadvantages of a decision-making process (Delgado-Gómez et al., 2019; Yannoukakou & Araka, 2014). DTA is used to identify and calculate the value of possible decision alternatives by considering the potential cost-adverse effects in the case of opening data.

A DTA for opening data (DTAOD) aims to estimate the costs and benefits of disclosing data. This will help us gain insight into the potential of using DTA to support data opening. A decision tree is a decision support tool that uses a tree-like model of decisions and possible consequences of conditional control statements (Yuanyuan, Derek, & Bob, 2018; Zhou & Wang, 2012). DTA is chosen as it can serve several purposes when complex problems in the decision-making process of disclosing data are encountered. Many complex issues in decision-making might be represented in the payoff table form (Song & Lu, 2015). Nevertheless, for the complicated problem related to costs and benefits analysis, DTA is very useful to show the routes and alternatives of the possible outcomes (Yuanyuan et al., 2018).

The DTA method consists of the following four steps (Adina Tofan, 2015; Delgado-Gómez et al., 2019): First, define a clear decision problem to narrow down the objective's scope. Factors relevant to alternative solutions should be determined. Second, structure the decision variables into a decision-tree model. Third, assign payoffs for each possible combination of alternatives and states. In this step, payoffs estimation is required to represent a specific currency of amount based on the experts' judgment. Fourth, provide a recommendation of decisions for the decision-makers. This research can support decision-makers and other related stakeholders such as business enablers and researchers better to understand the problem structure and variants of opening data.

5.3.1 Method of Decision Tree Analysis

The existing literature provides insight into the advantages of using DTA in the decision-making process. First, DTA can generate an understandable estimation process and is easy to interpret (Delgado-Gómez et al., 2019; Yeoa & Grant, 2018). Second, DTA can consider both continuous and categorical decision variables (Delgado-Gómez et al., 2019; Yuanyuan et al., 2018). Third, DTA provides a clear indication of which variable is becoming the most important in predicting the outcome of the alternative decisions (Adina Tofan, 2015). Fourth, a decision tree can perform a classification without requiring in-depth knowledge in computational (Delgado-Gómez et al., 2019; Song & Lu, 2015).

The DTA aims to manage several variables of the costs and benefits in opening data. Furthermore, the DTA can support the decision-makers in deciding how to select the most appropriate decision. Furthermore, this method can subdivide heavily skewed variables into specific ranges. In the DTA method, the decision-makers are trying to find the expected monetary value (EMV) of probability decisions, namely open dataset and limited access to the dataset. The EMV is the probability-weighted average of the outcomes (Delgado-Gómez et al., 2019; Yuanyuan et al., 2018). The use of EMV in DTA has two main benefits. First, EMV helps decision-makers to understand the possible investments of alternative actions. Second, DTA supports selecting the most appropriate alternatives by weighing the costs of two alternative decisions.

We used experts ' judgments to quantify the decision alternatives and possible paths to assign payoffs possible consequences of the costs and benefits in opening data, including the changes. The expert judgments were used because of their ability to interpret and integrate complex problems in a knowledge domain (Beaudrie, Kandlikar, & Ramachandran, 2016; Veen, Stoel. To do so, we interviewed four experts (three postgraduate researchers and one professional) in open government data and cost-benefits investments. The experts were selected based on their knowledge in the open data field.

The selected experts use their understanding and reasoning processes as they refer to their expertise and experiences for making judgments (Mach, Mastrandrea, Freeman, & Field, 2017; Walker, Catalano, Hammitt, & Evans, 2003). However, understanding the current issues and having logical reasons behind predicting costs and benefits in the open data domain is not trivial. The costs and benefits estimation requires sufficient knowledge and complex experiences in a specific field (Rush & Roy, 2001). There are some barriers and limitations of the expert judgment’s elicitation. First, during the elicitation process, the experts might quantify the answers not consistently because of the interviewer's unclear set of questions. To cover this issue, we design a protocol consisting of a list of questions to ensure a structure that was easy to comprehend by the experts. The use of specific terminologies in the field of open data, for instance, should be clearly defined. Second, the use of experts' judgment is potentially time-consuming, and experts are often overconfident that can lead to uncertainty estimation (Beaudrie et al., 2016; Knol, Slottje, Sluijs, & Lebret, 2010). We used aggregate quantitative review by subdividing heavily skewed variables into a specific amount of ranges to tackle this issue. Therefore, to convince the costs-benefits across the various experts as the stakeholders, we combined the elicitation results and the potential investments of opening data from the literature study.

5.3.2 Steps in developing the DTA

This study uses four main steps in developing DTA to manage and construct a decision tree-based analysis effectively and represent a schematic and structured way (Delgado-Gómez et al., 2019; Yeoa & Grant, 2018; Yuanyuan et al., 2018).

First, define a clear problem to narrow down the scope of the DTA. Relevant factors resulting in alternative solutions should be determined as well. This step involves both internal and external stakeholders to seek the possible options for a better decision-making process.

Second, define the structure of the decision variables and alternatives. The structure of the problems and influence diagram need to be described in a hierarchical model. In this step, organisations need to construct decision problems into tree-like diagrams and identify several possible action paths and alternatives.

Third, assign payoffs and possible consequences. In this step, the EMV formula is required to quantify and compare the costs and benefits. EMV is a quantitative approach that relies on specific numbers and quantities to estimate and calculate instead of using high-level approximation methods, such as agree, somewhat agree, and disagree options. For this, experts' judgment is used to estimate the payoff of possible consequences of the costs and benefits and estimate the chance of occurrence.

Fourth, provide alternative decisions and recommendations. After successfully assigning payoffs the possible consequences and considering adjustments for both costs and benefits, decision-makers can select the most appropriate decision that meets the success criteria and fits their budget. These steps will be followed when developed the DTAOD.

5.3.3 Analyse the dataset using DTA

Hereafter we illustrate the working of DTA by following the four steps.

Step 1: Define the Problem

The problem of opening data consists of three main aspects. First, decision-makers lack knowledge and understanding in estimating the costs and benefits of the open data domain and its consequences. Second, decision-makers do not know how to decide on the opening of data. Too much data might remain closed due to a lack of knowledge of alternatives. Third, decision-makers have no means to estimate the potential costs and benefits of opening data.

Step 2: Structure the Decision Alternatives

The decision-making process in opening data can be time-consuming and might require many resources. The decision-makers require simplifying complex and strategic challenges. Therefore, the DTA presented in this study can construct a model and structure the decision alternatives, whether the data should be released or closed.

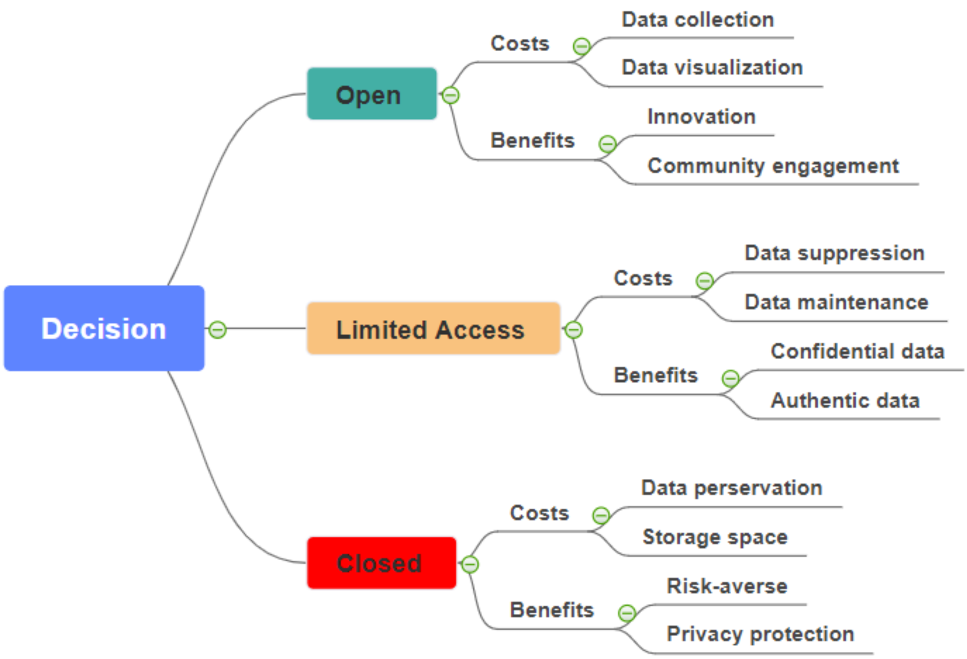


Figure 5-12 Decision alternatives and possible paths

Figure 5-12 illustrates the decision alternatives and various possible paths in deciding the complex problems of opening data. The nodes show that are three types of decisions, namely "open", "limited access", and "closed". The first decision refers to the governments releasing their data to the public with less or without restrictions. Second, limited access indicates that the access will be restricted to a specific group of users. Third, a closed decision refers to the government should keep the data exclusively.

Step 3: Assign Payoffs and Possible Consequences

In this step, numerical values to the probabilities are assigned, including the action-taking place and the investment value. In this study, the assigned payoffs represent the outcome for each combination in a table, namely a table of payoffs and possible consequences. This table uses costs terminology that describes the negative impact, such as value for the expense and potential lost revenue (Adina Tofan, 2015; Delgado-Gómez et al., 2019). At the same time, benefits-averse indicate the positive influence, such as a net revenue stream, potential income, and other profit elements (Adina Tofan, 2015; Song & Lu, 2015). The result of the assigned payoffs and the possible consequences are derived from the selected experts, as presented in Table 5-5.

Table 5-5 Assign payoffs and possible consequences

Alternative Decisions	Expert judgment (probability in percentage)					Expert judgment (invesment in Euro)						
	1	2	3	4	Mean	1	2	3	4	Mean	Total	Outcome
1. Open												
- Costs factors												
a. Data collection	65	67	58	62	63	15.500	16.200	16.500	16.600	16.200	30.238	46.438
b. Data visualization	35	33	42	38	37	14.250	15.500	12.100	14.300	14.038		44.276
- Benefits factors												
c. New knowledge	58	62	54	63	59	12.300	14.450	14.000	13.000	13.438	26.796	40.234
d. Community engagement	42	38	46	37	41	15.235	11.600	13.800	12.800	13.539		40.335
2. Limited Access												
- Costs factors												
e. Data supression	66	58	54	55	58	16.000	16.500	17.000	14.500	16.000	32.725	48.275
f. Data maintenance	34	42	46	45	42	16.000	17.000	16.800	17.100	16.725		49.450
- Benefits factors												
g. Confidential data	55	65	44	45	52	18.000	17.600	17.700	18.200	17.875	35.000	52.875
h. Authentication data	45	35	56	55	48	18.500	17.500	16.850	16.500	17.338		52.338
3. Closed												
- Costs factors												
i. Data preservation	72	68	62	70	68	13.000	14.500	13.500	14.200	13.200	27.588	40.788
j. Storage space	28	32	38	30	32	16.000	15.850	12.200	13.500	14.388		41.976
- Benefits factors												
k. Risks-averse	52	56	57	60	56	9.300	10.500	12.000	10.000	10.450	22.513	32.963
l. Privacy protection	48	44	43	40	44	11.000	13.000	11.750	12.500	12.063		34.576

Table 5-5 presents the result of the assigned payoffs between three alternative decisions, namely: "open", "limited access", and "closed". This table includes the expert judgment in estimating the probabilities of the costs and benefits and the numerical values given to predict the investment of money in the euro currency. When the entire process of assigning payoffs has been completed, we can calculate the average numerical values of the costs and benefits percentages possibilities. For example, data collection might invest 63% of the revenue stream instead of a data visualisation program (37%). This means that the most significant money investment from this opening decision is data collection.

Data collection refers to a mechanism of gathering the dataset on the variables of interest from the holders or owners by using specific manners and techniques (S. Kim & Chung, 2019). Furthermore, data visualisation refers to presenting the dataset into an interactive and user-friendly interface and effectively capturing the essence of the data (Xyntarakis & Antoniou, 2019). Regarding the potential investment of money between data collection and data visualisation, Figure 5-13 shows that deriving data from data providers can have higher expenses than visualising the data. In addition, according to experts, data collection requires more than 16,000 Euros on average of investments, which is higher than data visualisation about (14,000 Euros). Therefore, the total costs for opening data collection and visualisation data decisions equal approximately 30,000 Euros. Figure 5-13 shows the complete decision tree showing all alternatives.

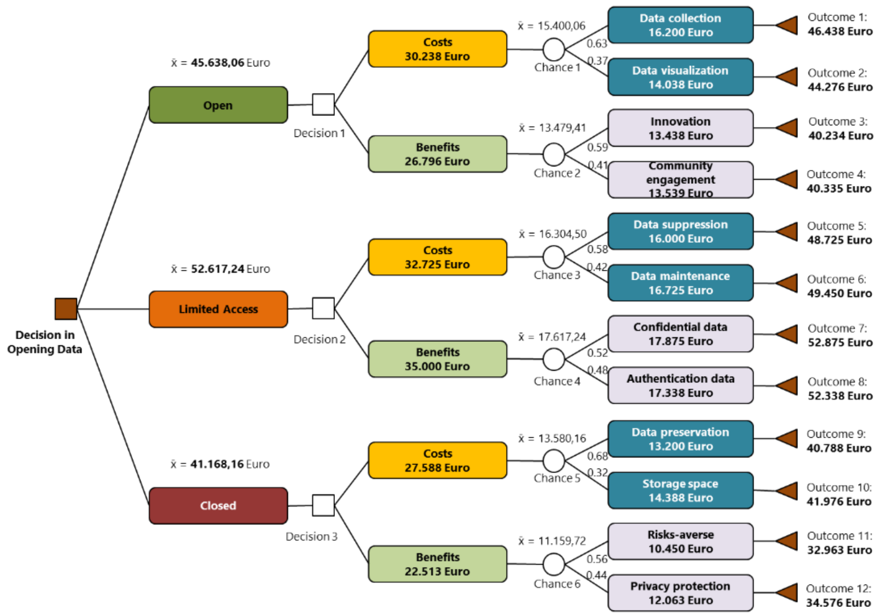


Figure 5-13 Decision tree analysis to estimate the costs and benefits

The decision tree presented in Figure 5-13 results in the payoff result depicted in Table 5-5. From the constructed data, we can compare the costs and benefits of the three decision nodes. The values stated on each sub-element are the prediction of monetary expenses. For example, we have to do some structured ways to obtain the expected monetary value from an open decision. First, we need to know the average costs of data collection and data visualisation by calculating the amount's probability and estimation. Here, we calculate (0,63 x 16.200 Euro) + (0.37 x 14,038 Euro) = 15,400.06 Euro. Second, we estimated the open data decision costs by adding up the value of data collection and data visualisation (16.200 Euro + 14.038 Euro) = 30.238 Euro. Third, the experts estimated the outcome for each sub-cost and sub-benefit factor. Next, data collection and visualisation should be added to the potential total costs (16.200 Euro + 30.238 Euro) = 46.438 Euro (outcome 1). Whereas outcome two is obtained from (14.038 Euro + 30.238 Euro) = 44.276 Euro. After doing this in the same way for the benefits factors, we estimated the open decision's total investment. Before calculating the sub-costs and sub-benefits, it is important to compare the highest potential investment between the costs and benefits factors. The reason is to determine the highest priority of the potential investment between costs and benefits consideration. In this case, the highest probability is the cost factors (30.238 Euro) instead of its benefits (26.796 Euros). Therefore, the total average expected monetary value (EMV) for "open" decision is equal to the EMV of the costs adding up to the total value of the costs whereby 15.400,06 Euro + 30.238 Euro = 45.638,06 Euro.

Step 4: Provide decisions and recommendations

Based on the constructed decision tree (in Figure 5-13), the final step in the decision tree analysis is making a decision and providing some recommendations presented in decision action plans. We weigh the open data costs and benefits to the decision-makers to provide the most suitable decision between the three alternatives (open, limited access, and closed). Next, from the EMV results, the DTA can recommend a decision based on the biggest influence on investment in institutional revenue streams. We classify the findings of the study into two parts, namely:

1. Possible paths and with total payoffs

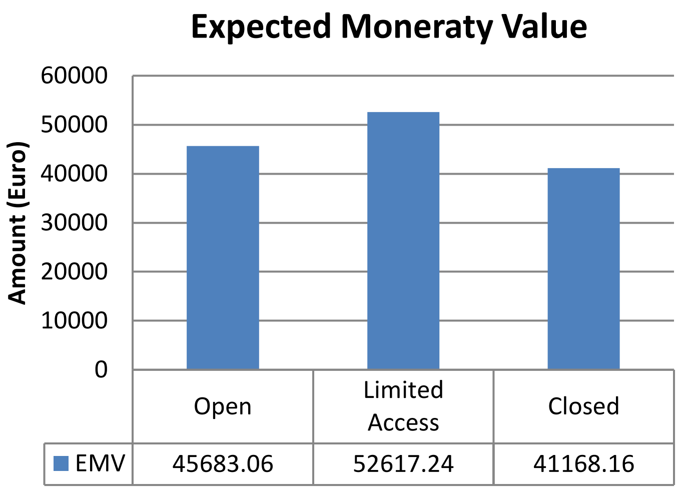
The first finding from the decision tree analysis is the possibility of the nodes and paths and their chances, as can be seen in Table 5-6. Every decision alternative provides the estimation of payoffs in the Euro currency. Based on these results, it can be concluded that the highest investment for the costs factor in the open data domain is data maintenance, where the cost is almost 50,000 Euros. Data maintenance is a sub-node of the limited access decision. The highest potential benefit by implementing the decision is the confidentiality of the data, where about 52,000 Euros would benefit the government institutions. In this case, the limited access decision can potentially have high costs, resulting in high social benefits.

Table 5-6 Possible nodes, paths, and estimation payoffs

Terminal	Total Payoff
Decision → Open → Decision 1 → Costs → Chance 1 → Data collection	46.438 Euro
Decision → Open → Decision 1 → Costs → Chance 1 → Data visualisation	44.276 Euro
Decision → Open → Decision 1 → Benefits → Chance 2 → New knowledge	40.234 Euro
Decision → Open → Decision 1 → Benefits → Chance 2 → Community engagement	40.335 Euro
Decision → Limited access → Decision 2 → Costs → Chance 3 → Data suppression	48.725 Euro
Decision → Limited access → Decision 2 → Costs → Chance 3 → Data maintenance	49.450 Euro
Decision → Limited access → Decision 2 → Benefits → Chance 4 → Confidential data	52.875 Euro
Decision → Limited access → Decision 2 → Benefits → Chance 4 → Authentication data	52.338 Euro
Decision → Closed → Decision 3 → Costs → Chance 5 → Data preservation	40.788 Euro
Decision → Closed → Decision 3 → Costs → Chance 5 → Storage space	41.976 Euro
Decision → Closed → Decision 3 → Benefits → Chance 6 → Risks-averse	32.963 Euro
Decision → Closed → Decision 3 → Benefits → Chance 6 → Privacy protection	34.576 Euro

2. Expected monetary value (EMV)

The expected monetary value (EMV) resulting from the DTA shows that the limited access decision could gain about 52,000 Euro's highest monetary value for society. It follows the open decision in approximately 45,000 Euros, and the decision to keep closed the data can contribute around 41,000 Euros. The EMV of each decision is derived from the probability-weighted average of the expected outcome. Figure 5-14 presents the detailed EMV result and investment ranges. This EMV result can help decision-makers in estimating and quantifying the investments needed.



Decision	EMV	Range
Open	45.638,06 Euro	44.276 – 46.438 Euro
Limited Access	52.617,24 Euro	52.338 – 52.875 Euro
Closed	41.168,16 Euro	40.788 – 41.976 Euro

Figure 5-14 The expected monetary value and investment ranges

In summary, many government organisations are reluctant to disclose their data because they have limited insight into the potential costs and possible adverse effects. Processing data or opening datasets partly can overcome this problem. However, this requires investments. This study presented the DTAOD method to estimate the potential investments and merits of opening a dataset. There are several advantages found in using DTAOD in this study. First, the decision tree can provide a better understanding of the possible outcomes of a decision alternative. Second, the proposed decision tree provides insight into selecting an informed decision. However, this is highly dependent on the alternatives that are formulated and included in the decision tree. Third, the decision tree can allocate the values in estimating the costs and benefits in the open data domain based on expert judgments. This provides insight into the activities needed for opening data and the associated costs and benefits.

At the same time, using DTAOD might not be easy. First, during the assigned payoff process, a small change in the quantification of numerical values can significantly change the decision tree's entire structure. Second, the calculations are based on experts' information, but these might not be correct or biased towards openness or closeness. This result shows that the high and low expected monetary values (EMV) of a decision will influence the decision made.

A DTA study contributes to a better understanding of the problem structure and provides new insight in estimating the costs and benefits of releasing data for the policy-makers. In future research, we recommend using a different method like paired comparison, multi-voting, and net present value (NPV) methods to quantify the assigned payoffs as this study using a single expert judgment.

5.4 A comparison of methods based on the literature

The disclosure of public sector information through open government data initiatives can provide numerous advantages to the public domain at a large scale (Zuiderwijk & Janssen, 2013b, 2015). Opening various datasets might drive high demand from stakeholders (Ubaldi, 2013; Veenstra & Broek, 2013). Analysing the advantages and disadvantages and other effects-adverse of disclosing data to the potential stakeholders is cumbersome (Luthfi, Janssen, et al., 2018a). Methods like BbN, FMCDM, and DTA, can be used to analyse the potential advantages and disadvantages of opening data (Ali-Eldin et al., 2017; Luthfi, Janssen, et al., 2018a; Luthfi, Rehena, et al., 2018). However, it is unclear which one is more suitable. Therefore, there is a need to compare them. We use systematic literature as the main sources to compare the methods in this section. We first identify the criteria to compare the methods and thereafter, we compare the methods.

5.4.1 Comparison criteria

The comparative method comprises feature-by-feature of selected parameters (Jan, Farman, Khan, Imran, & Islam, 2019; Nojava, Qian, & Stow, 2017). In this study, the comparison parameters will be divided into three main parts. First, the input parameter consists of three criteria: experimental data, data types, and posterior probability (Nojava et al., 2017). Second, the process parameters are decomposed into four criteria: efficiency, easiness, effectiveness, and complexity. Third, the output parameters are comprised of three criteria: transparency, subjectivity, accuracy, and usability. The three sub-parameters used can be explained in detail, as follows:

1. **Input**

   The input data is different for each of the methods. For comparing the methods, experimental data, data types, and posterior probability are relevant (Stutz & Cheesman, 1994). First, experimental data refers to data produced in measurable activities by doing an experimental or quasi-experimental design (Beuzen, 2018). In our illustrations, the experimental data originates from expert input. The experimental data may be quantitative or qualitative using different investigation methods. Second, dataset type (CSV format) refers to a specific type of dataset presented in tabular form. Each column of the table represents a specific meaning of values (Safarov et al., 2017). Third, posterior probabilities define an uncertainty proposition of the conditional probability allocated after the relevant evidence is considered (Luthfi, Janssen, et al., 2018a).
2. **Process**

   The process can be viewed from 4 criteria, e.g., efficiency, easiness, effectiveness, and complexity. Efficiency refers to avoiding wasting efforts, energy, and time in the evaluation process (Nkurunziza, 2019). It is a measurable instrument of the selected variable in a mathematical sense to ensure the effort to produce and establish a specific outcome with minimum costs and endeavour (Beuzen, 2018). Second, the easiness of the selected method in analysing the selected method means the ease of manner and evaluation process rules (Chakraborty et al., 2016).

Third, effectiveness refers to the capability of generating the desired result, which means it has an expected outcome and a clear impression (Nojava et al., 2017). Fourth, the process's complexity refers to a system's behaviour in interacting components in multiple ways and reasonable (Beuzen, 2018).

3. **Output**

The output parameters are made up of three criteria, e.g., transparency, subjectivity, accuracy, and usability. The first sub-parameter considers the transparency of the process. Transparency means that the evaluation process is easy to recognise and be understood (Mallach, 1994). Second, subjectivity refers to a subject's insights and judgments influenced by personal feelings, desires, expertise in discovering, and level of beliefs in terms of phenomena (Beuzen, 2018). Third, the accuracy of the results in evaluating means the accuracy and precision of measurements (Chakraborty et al., 2016). A measurement system in specific could be accurate but not precise and vice versa. Fourth, usability is defined as the capacity of the proposed system or tool to provide a condition for a specific user within a particular context and interface to perform the tasks (Nielsen, 2012).

5.4.2 Comparison of the methods

The following Table 5-7 summarises our comparison of the three methods for open data decision-making. The comparison is based on reviewing each method from the literature and applying them using an illustrative example. This table confirms our starting point for selecting the methods. The methods have different characteristics, and they might yield different benefits when used for deciding to open data.

Table 5-7 Comparative the methods in opening data

Parameter	Bayesian-belief Networks	Fuzzy Multi-criteria Decision Making	Decision Tree Analysis
Input			
Experimental data (from experts)	Data is summarised based on the likelihood function from the observed	Data is summarised based on the pairwise comparison	Data is summarised based on the assigned payoffs process of possible

Parameter	Bayesian-belief Networks	Fuzzy Multi-criteria Decision Making	Decision Tree Analysis
	dataset (Luthfi, Janssen, et al., 2018a; Nojava et al., 2017)	matrix (Luthfi, Rehena, et al., 2018)	investments (Delgado-Gómez et al., 2019)
Data type	Numerical and categorical (Herland et al., 2016)	Numerical and categorical (Ceballos et al., 2017)	Numerical and categorical (Yeoa & Grant, 2018)
Probability	Posterior probability distribution (Nojava et al., 2017)	Posterior probability distribution (Ceballos et al., 2017)	Posterior and conditional probability distribution (Adina Tofan, 2015)
Process			
Efficiency	Time-consuming (maximum) (Herland et al., 2016)	Time-consuming (moderate) (Kahraman et al., 2015)	Time-consuming (minimum) (Yuanyuan et al., 2018)
Easiness	Highly difficult to understand and interpret the model. Advanced in the mathematical background is required (Horný, 2014)	Moderately difficult to understand and interpret the model. Advanced in the mathematical background is required (Mohsen et al., 2014)	Relatively easy to understand and interpret the model. The basic mathematical background is required (Yuanyuan et al., 2018)
Effectiveness	Constructing a causal relationship between variables and provide decision recommendations (Beuzen, 2018)	Constructing a hierarchy of decisions including its alternative and ranking them into best options (Rezaei et al., 2013)	Constructing a structured decisions estimation and its consequences (Adina Tofan, 2015)
Complexity	Require the size of the belief network to simulate and construct complex conditional probabilities (Luthfi, Janssen, et al., 2018a)	Require rule base analysis to construct a pairwise comparison matrix (Rezaei et al., 2013)	Changing variables during the analysis process might be possible to redraw the existing tree. Irrational expectations can lead to flaws and errors in the decision tree (Yeoa & Grant, 2018)
Output			
Transparency	Require high level to comprehend the process and expected results	Require high level to comprehend the process and expected results (Werro, 2015)	Require a moderate level to comprehend the process and expected results

Parameter	Bayesian-belief Networks	Fuzzy Multi-criteria Decision Making	Decision Tree Analysis
	(Chakraborty et al., 2016)		(Delgado-Gómez et al., 2019)
Subjectivity	The elicitation data and information from the experts might possible bias the quantification process (Beuzen, 2018)	The elicitation data and information from the experts might somewhat bias the quantification process (Ceballos et al., 2017)	The elicitation data and information from the experts might possible bias the quantification process (Delgado-Gómez et al., 2019)
Accuracy	The expected value is more accurate when there is less uncertainty in the input parameter. The output is distributed over a range of uncertainties (Herland et al., 2016; Luthfi, Janssen, et al., 2018a)	The estimation result is more consistent compared to the reference data approach (Werro, 2015)	The expected result is accurate and able to predict the future outcome (Delgado-Gómez et al., 2019)
Usability	Moderately difficult to accomplish the objective of the BbN method in terms of time allocation. However, users can follow and learn the BbN method because of the detailed systematic steps provided in the system interface (Husmeir, 2005; Luthfi, Janssen, et al., 2018a)	Highly difficult to accomplish the objective of the FMCDM method. Users need extra time to learn and understand the detailed steps with the mathematical formulation and fuzzy linguistic matrix (Luthfi, Rehena, et al., 2018; Mohsen et al., 2014)	Relatively easy to accomplish the objective of the DTA method. Users do not spend a lot of time to learn and understand the detailed steps, including the construction of hierarchical decision alternatives (Luthfi et al., 2019; Mittal & Khanduja, 2017)

Table 5-7 shows the differences and similarities of the three methods. BbN requires maximum allocation time in processing the evaluation compared to the other two methods (Herland et al., 2016). This approach is noticeably challenging to understand and interpret the proposed model. Subjective judgments are needed in all three methods. The decision-makers of dataset officers require the capability in mathematics background (Horný, 2014). However, the advantage of using this method is the result is more accurate in terms of uncertainties (Chakraborty et al., 2016; Herland et al., 2016).

FMCDM requires fewer resources than BbN for evaluating the dataset (Kahraman et al., 2015). This method is relatively difficult to comprehend, and model interpret is not easy (Mohsen et al., 2014). The pairwise comparison tasks may also need an advanced understanding of mathematics (Rezaei et al., 2013; Werro, 2015). The expected results show moderate bias in the quantification process (Ceballos et al., 2017). The FMCDM's benefit is the dataset consistently estimates the selected parameter (Werro, 2015).

DTA is summarised based on assign payoffs the number of possible investment values (Delgado-Gómez et al., 2019; Yeoa & Grant, 2018). In this method, when using a manual process, it requires more time to re-construct the decision structures and paths when changing variables during the analysis process, and it might be possible to redraw the existing tree (Adina Tofan, 2015; Yeoa & Grant, 2018). However, the advantage of using DTA is that it is relatively easy to understand, and the model is relatively easy to interpret (Yuanyuan et al., 2018).

5.5 Conclusions

The three methods discussed were selected based on their diversity and different contribution to open data decision-making. The literature comparison shows that these methods have different pros and cons. The BbN requires more time in processing the evaluation in comparison with the other two methods. The advantage of using BbN method is the result more accurate and is able to deal with uncertainties. BbN can weigh the benefits and disadvantages of the opening dataset by taking into account uncertainties and conditional dependencies. FMCDM consumes less time than BbN in evaluating the dataset. The FMCDM method is relatively difficult to comprehend, and the model interpretation is not easy. Like BbN, it also needs an understanding of mathematics. The benefit of using FMCDM is the ability to systematically construct a hierarchy of decisions, including its alternative and ranking them into the best options. FMCDM can capture the expertise of open experts and express it with a computational approach based on intuitive reasoning by considering human subjectivity. DTA is the least time-consuming method; however, the DTA is challenging, and the re-construction of the decision structures is time-consuming when decision-makers modify variables during the analysis process. Furthermore, it might not be easy to redraw the existing tree. This can result in a lengthy process. Nevertheless, the advantage of using DTA is relatively easy to understand and interpret the model. As no historical data is available, we had to rely on the expert's input for all three methods, which is subjective as their knowledge and experience are limited. We suggest collecting historical data based on the use of these methods in further research. Chapter 7 will compare the three methods by experimenting with the opening of data by decision-makers. For decision-making, a prototype is needed, which will be discussed in the next chapter.

Chapter 6
Development of the DSOD prototype

This chapter discusses the third phase of this research, namely, the development of the DSDOD prototype. To develop the DSOD prototype, we followed a prototyping developing approach. This chapter is decomposed into six sections. First, we introduce the prototyping approach used in this research. Second, we present the prototyping objectives to develop detailed functionalities of the DSOD prototype. Third, we present the functions of the DSOD prototype based on the decision-making support elements. Fourth, we discuss the DSOD prototype's construction by following the decision-making process's conceptual model in Chapter 4. Fifth, we validate the DSOD prototype. Finally, we draw conclusions to answer the third research question.

Parts of this chapter have been published in (Luthfi & Janssen, 2017) and (Luthfi, Rukanova, Molenhuis, Janssen, & Tan, 2020).

6.1 Prototyping approach

The prototyping approach contains the steps of how the prototype was developed. Prototyping often follows four steps, namely (1) functional selection, (2) construction, (3) evaluation, and (4) further use (Floyd, 1984).

First, functional selection refers to the options of functions that the prototype should demonstrate. The selection mechanism should always be based on relevant tasks that can serve as model cases to exhibit the proposed prototype model. The diverse between the functional range of the prototype and the product may be that the system functions employed are offered in their intended final form. This situation is also known as the selected function from the vertical prototyping. However, if the prototype features are impossible to implement in detail, the (software) developers can simulate the prototype using horizontal prototyping.

Second, construction represents the means to make the prototype exist. This work should be much smaller than when developing the entire product. An appropriate functional selection and preferable techniques prototype is an important consideration to construct a prototype. During the prototype generation, the attention should be on the expected evaluation rather than taking account of the regular and long-term use. As a result, the construction of a prototype can fulfil the product's standardisation and quality requirements, like efficiency, time scheduling, and security impacts are all covered by this step.

Third, evaluation refers to collecting feedback for the further process of prototype development. Ensure the participation of all relevant groups of users in gaining feedback is needed. A training mechanism is needed before doing the actual evaluation steps.

Fourth, the prototype's further use refers to possibilities to develop the current prototype based on the experiences gained with the prototype and the available production environment. During the decision to create additional functionalities, the project organisation can fully or partially create a new module or function of the prototype to reach the expected prototype's target.

In this research, we adapted to the four steps (Floyd, 1984) and added a new step, "defining prototype objective," to establish and filter the prototype functionalities and user requirements. The five steps decompose into (1) define prototype objective, (2) define the prototype approach, (3) selecting the functions and elements of the DSOD prototype, (4) constructing the DSOD prototype, and (5) determine the testing mechanism of the DSOD prototype. These steps focus on using the DSOD prototype to evaluate the advantages and disadvantages of opening data. The detailed steps of the prototyping approach of this study are outlined in Figure 6-1.

Step 1. Define the prototype objectives

- To define and filter the prototype functionalities and user requirements.
- To evaluate the outcome of the DSOD prototype functionality from the group of participant's experiments.

Step 2. Define the prototype approach

In this research, we used evolutionary prototyping approach to explore the strengths and weaknesses of the DSOD prototype. There are six main evolutionary process used in this step, as follows:
(1) Requirement gathering of the prototype
(2) Quick design of the prototype
(3) Develop a prototype
(4) Evaluate the prototype based on user's feedback
(5) Refine the prototype
(6) Re-construction the prototype

Step 3. Selecting the functionalities and elements of the DSOD prototype

We used the elements and functionalities of the DSOD prototype, which are consists of six main processes.
(1) Retrieving and decomposing the dataset
(2) Analyzing the dataset
(3) Weighing the dataset
(4) Decision-making alternatives
(5) Updating the status of the dataset

Step 4. Constructing the DSOD prototype

We developed a mock-up based prototyping to simulate the decision-making process to open data. The mock-up contains the elements and functionalities defined in the step 3.

Step 5. Determining the prototyping testing mechanism

To evaluate the DSOD prototype, we used alpha test and beta test. The alpha test aims to discover apparent problems and issues before a first version of the prototype released. The objective of the beta test is to look at the user's working environment.

Figure 6-1 The prototyping approach and steps in this research

In step 1, we defined the prototype's objectives. The first objective is to establish and filter the prototype functionalities and user requirements. The functionalities of the prototype are derived from the elements of the DSOD prototype in section 3.3.5. We defined the user requirements by detailing the specification and levels of the users. The second objective of the prototype is to evaluate the outcome and effects of the DSOD prototype functionality from the participant's experiments.

In step 2, we used the evolutionary prototyping approach (Bernstein, 1996). The selection of the evolutionary approach in this study begins with the initial feedback from ten OGD stakeholders consisting of three decision-makers, three civil servants, two data analysts, and two PhD students. We added additional functionalities until the stable version is released. The evolutionary prototyping approach differs from the rapid prototyping scheme. The evolutionary prototype starts with a better understanding of the stakeholders' decision-making elements and requirements from the literature and feedback. In rapid prototyping, the developer implements the least understood requirements (Koutsoukis et al., 2000; Matthews & Wensveen, 2015). There are six main steps to process the evolutionary prototyping (Bernstein, 1996; Budde, Kautz, Kuhlenkamp, & Züllighoven, 1992; Floyd, 1984), namely (1) requirement gathering of the prototype, (2) quick design of the prototype, (3) developing a prototype, (4) Evaluating the prototype based on user's feedback, (5) refining the prototype, and (6) re-constructing the prototype. One of the advantages of using evolutionary prototyping in this research is that it's useful for exploring intelligent systems of decision-making support of opening data.

In step 3, we selected the functionalities and elements of the DSOD prototype. In this step, we used the elements and functionalities of the DSOD prototype, which consists of five main processes. The five main elements include: (1) retrieve and decompose datasets, (2) analysing the dataset, (3) weighing the dataset, (4) decision-making alternatives, and (5) updating the status of the dataset.

In step 4, we constructed the DSOD prototype using a mock-up to simulate the decision-making process to open data. The mock-up- objective in this research is to test the design and functionalities of the DSOD prototype (Koutsoukis et al., 2000).

Furthermore, the mock-up can point to the clashing visual elements of our DSOD design and help organise the visual features in detail. This step can determine the necessary revisions to be carried out into depth system requirements, such as user interface contrasts, colours, and visual hierarchy.

Finally, to evaluate the DSOD prototype, we use the alpha test and beta test in step 5. The alpha test aims to discover apparent problems and issues before a first version of the prototype is released and tested to the participants (Sommerville, 2011). The alpha test aided in investigating unexpected system behaviour during the evaluation of the DSOD prototype. The beta helped to explore the interaction problems and issues regarding the developed prototype and the environment. Moreover, the beta test helped ensure that the actual users or participants can complete their tasks and get a wide range of user interactions (Beaudouin-Lafon & Mackay, 2002; Sommerville, 2011).

6.2 Prototype functionality

This research designed the functionalities and the DSOD prototype's selection into five main functionalities, as we described in Chapter 4. The five functionalities include (1) Initialisation, (2) analysing the dataset, (3) assessing the dataset, (4) decision-making alternatives, and (5) updating the status of the dataset. The detailed steps of the decision-making process prototype, as follows:

1. ***Initialisation***

   We used the extract, transform, and load (ETL) approach to retrieve and decompose the selected dataset in this step. The ETL approach aims to process a dataset chosen that involves collecting the dataset from external data providers and converting it into a machine-readable format (Gour, Sarangdevot, Singh, & Sharma, 2010; Vassiliadis & Simitsis, 2009). The extraction step is a process to read and identify the correct subset of source data that has to be submitted to the ETL workflow for further processing steps (Vassiliadis & Simitsis, 2009). In the context of DSOD, the DSOD system will filter the selected table of the dataset, including ensuring that all fields are intact and maintained concerning each table.

A variety of datasets structure formats can be read by the system both based on proprietary database or open platforms such as ACCDB (Microsoft), CDF (XML standard), Comma-separated Value (CSV), Database File (DBF), Digital Elevation Model (DEM), ESRI (Geo DB), JavaScript Object Notation (JSON), and others.

Subsequently, the DSOD system transforms the datasets to convert the extracted datasets from their previous form into the database format (e.g., CSV, XML, and JSON). A dataset's transformation is required in the decomposing process by using specific rules or lookup tables, like combining attributes of a dataset with another relevant field in a table (Gour et al., 2010). In the data transformation step, a series of functions in the DSOD system is responsible for preparing the extracted data into the end of the targeted database. One of the critical functions, namely data cleansing, which aims to pass only the structured data, can be restored to the targeted database.

Finally, the DSOD system loads the dataset to record the datasets into the targeted database in this step. There are several objectives of the loading process in this step. First, the DSOD ensures that the selected key data field has neither missing values nor null. The DSOD will check whether the value of each attribute is in the proper types of data. Second, the DSOD tests the selected dataset into modelling views based on the targeted tables. In this process, the DSOD specifies the join statements between the selected attributes in a table to view and present the data into a single table. In this way, the prototype can help to analyse the dataset. Third, the DSOD checks the combined values of the dataset in the appropriate format to ensure that dimension tables have a data dictionary and history of the table. This step aims to help the DSOD users trace the historical process of the extraction and transformation steps.

2. ***Analysing***

This stage is a critical stage where the datasets that have been selected in the previous step will go through the evaluation process. The system will interpret data that translates each data value from a table to be included in two broad categories of advantages and disadvantages. Datasets are evaluated using the three selected methods (see Chapter 5), namely BbN, FMCDM, and DTA, whereby every method has a different number of sub-steps. In the BbN method, there are five main sub-steps to analyse the selected dataset. First, the DSOD determines each selected attribute's possible advantages and disadvantages factors. Second, the DSOD constructs a causality network structure to show each attribute's cause and effect in terms of the advantages and disadvantages factors. Third, the DSOD formalizes the structure of the causality network to subjective prior beliefs from experts. Fourth, the DSOD quantifies the prior probability of defining and estimating the probability distributions for each benefit and cost factor. Fifth, the DSOD interrogates the belief network to set the sensitivity level of the advantages and disadvantages factors.

In the FMCDM, there are four main sub-steps to analyse the selected dataset. First, the DSOD determines the possible criteria and sub-criteria of each attribute's advantages and disadvantages factors. In this step, the user can select the potential advantages and disadvantages criteria based on their knowledge and expertise. Second, the DSOD constructs a relationship diagram to show the relationship between criteria and sub-criteria of the advantages and disadvantages factors. Third, the DSOD constructs the hierarchical diagram to show the criteria, sub-criteria, and decision alternatives. Fourth, the DSOD defines decision alternatives for each criterion and sub-criteria.

In the DTA, there are two main steps to analyse the dataset. First, the DSOD determines the advantages and disadvantages. In this step, the user can select the possible benefits and disadvantages criteria based on their knowledge and expertise. Second, the DSOD constructs a decision structure to show the tree structure of alternative decisions and their advantages and disadvantages.

3. ***Weighing***

The previous stages' evaluation results are classification and level references to the advantages and disadvantages of datasets. This system's advantage is to provide iterative process conditions to ensure that the benefits level is higher than the disadvantage at hand. Technically, during the assessment process (step 3 in Figure 6-1), the system will combine the overall scores from the benefit and disadvantage analyses to determine the appropriate solution for treating the dataset.

In the BbN method, four sub-steps weigh the potential advantages and disadvantages of the previous steps (analysing). First, DSOD needs to evaluate the information in the dataset and the advantages and disadvantages factors. Second, DSOD should observe the potential benefits and users of the dataset against the likelihood of their evidence. Third, DSOD evaluates the potential costs to look at the disadvantages adverse effects of the dataset against the likelihood of their evidence. Fourth, DSOD will process the weigh mechanism by integrating the overall quantified result from the previous step.

In the FMCDM method, there are three main steps to weigh the potential advantages and disadvantages factors. First, we initialised the process of selecting and extracting the medical records dataset. Next, we defined the four criteria and another eight sub-criteria. Second, we evaluated the criteria and sub-criteria of the potential advantages and disadvantages of opening data by using three steps: developing a linguistics matrix, employing fuzzy AHP, and quantifying the fuzzy linguistics scale using an expert's judgment. Third, we provided alternative decisions.

At the same time, there are two main steps in the DTA method to weigh the potential advantages and disadvantages factors. First, the DSOD assigns the decision table's payoffs to describe the negative impacts of decisions like value for the expense and potential lost revenue. Second, the DSOD constructs the possible consequence, including the action-taking place and the investment value expected as the outcome.

4. ***Decision-making alternatives.*** There are four possible decisions to release the datasets. *Open decision* refers to publishing the dataset without any additional measures. This suggests that the disadvantages are low that and the benefits of releasing the dataset substantially outweigh the estimated disadvantages.

Limited access decision refers to the releasing dataset will have moderate disadvantages, or the dataset's potential benefits do not outweigh the potential disadvantages. *Additional screening decision* refers to publishing the dataset, it tends to create significant disadvantages, and the potential benefits do not outweigh the potential estimated disadvantages. Finally, a *closed decision* to keep the dataset closed suggests that releasing the dataset generates a disadvantage outweighing the potential benefits according to the decision-makers.

5. ***Update the status of the dataset.***

   In this step, when a dataset is decided to open with limited access, the DSOD can iterate the process to keep higher benefits than the costs by updating the dataset and going back to the analysing process in the second step. The DSOD supports the iterative process to achieve of consolidated solution between the users and decision-makers. The decision-makers can use the DSS to modify, refine, and evaluate the selected dataset before sending it back to the cycle process.

6.3 Prototype construction and demonstration

This section presents the prototype's construction by following the decision-making process's conceptual model in Chapter 4. Also, each of the steps is presented. Figure 6.2 shows the navigation system of the DSOD prototype. The prototype's navigation aims to provide a structural way that follows the decision-making process functionalities to open data. The five functionalities include initialisation, analysing the dataset, assessing the dataset, decision-making alternatives, and updating the dataset's status.

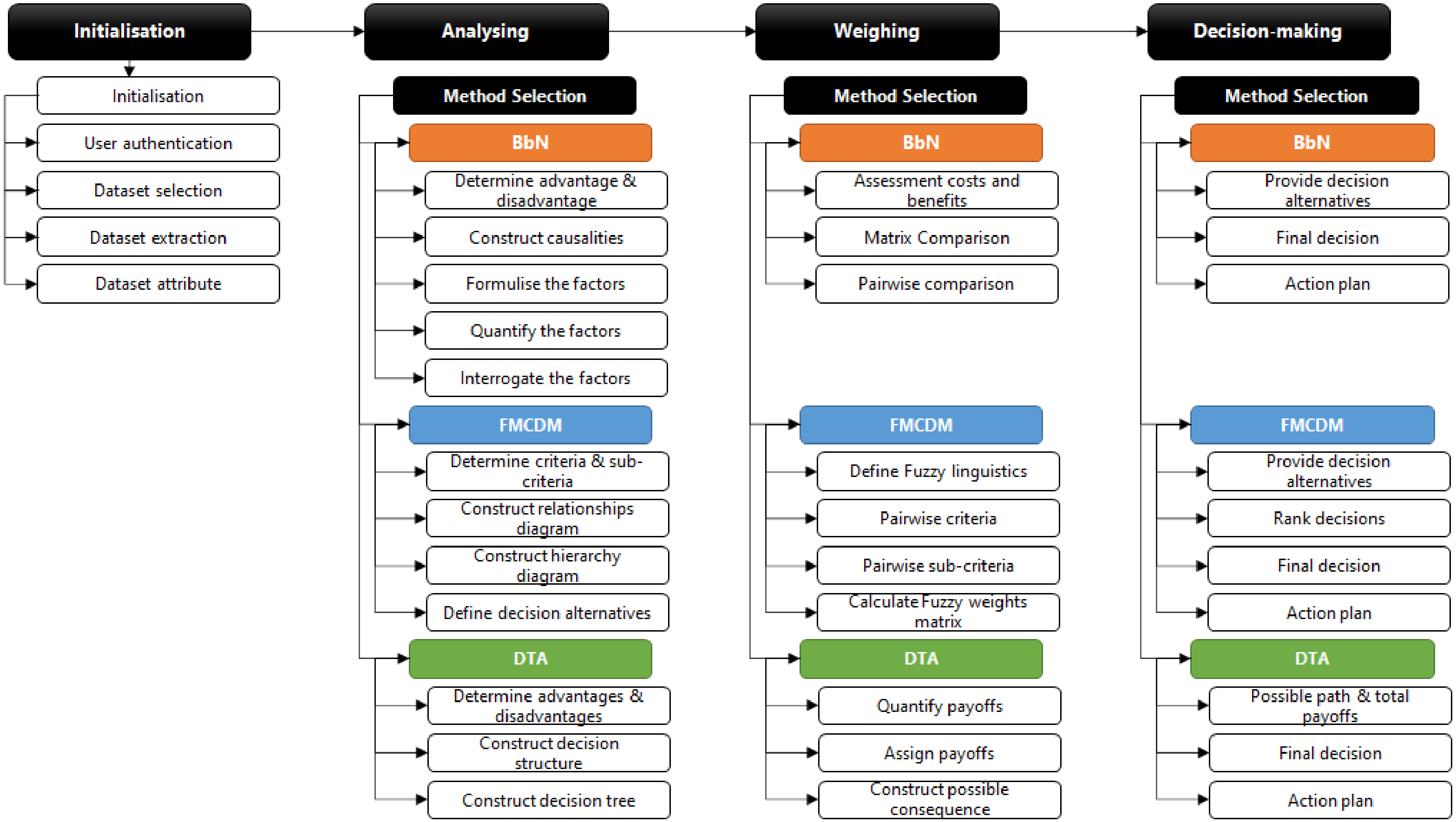


Figure 6-2 The navigation systems of the DSOD prototype

6.4.1 Initialisation

In this process, we used five main sub-steps, namely (1) user authentication, (2) message box, (3) dataset selection, (4) dataset extraction (metadata), and (5) dataset attribute. These five sub-processes are based on the DSDOD functionalities defined in Chapter 3 and Chapter 4. First, the user authentication process indicates the users' groups and levels: administrator, data analyst (experts), and decision-makers. The DSOD provides access and privilege for the users to analyse and decide alternatives in the DSOD prototype. Second, the message box is used to notify the users what the current status of the previous process of analysing a dataset is.

Furthermore, the DSOD will send recent instructions or tasks to the users. For example, the regular schedule to examine the new dataset or reiterate and update the previous datasets. Therefore, this feature could help users be aware of their tasks and responsibility in the DSOD prototype. Third, the data selection process aims to select the dataset, including its category, format, and destination of the dataset source.

Fourth, the DSOD provides the extraction process to retrieve and decompose the selected datasets in the data extraction process. The extraction step aims to read and identify the correct subset of source data that has to be submitted to the extraction workflow for further processing steps. In the context of DSOD, the DSOD system will filter the selected table in the dataset, including ensuring that all fields are intact and maintained concerning each table. A variety of datasets structure formats can be read by the system both based on proprietary database or open platforms such as ACCDB (Microsoft), CDF (XML standard), Comma-separated Value (CSV), Database File (DBF), Digital Elevation Model (DEM), ESRI (Geo DB), JavaScript Object Notation (JSON), and others. Subsequently, the DSOD system transforms the datasets to convert the extracted datasets from their original form into the standardised format to enable processing.

Fifth, the last sub-step of the initialisation process is that the DSOD will show the dataset attributes. There are two types of attribute information provided in this step, e.g., attribute name and data type. The attribute name indicates the original attribute names of the selected dataset. The data type can be shown in several formats, such as variable character (varchar), currency, and an integer.

Figure 6-3 The main interface of the DSOD prototype

6.4.1.1 User authentication

Authentication is the process of determining whether a user is, who, or what it declares itself to be recognised (Aldossary & Allen, 2016). We used user authentication based on the DSOD elements and functionalities in Chapter 3. An authentication interface provides access control for systems by checking to see if a user's credentials match the credentials in a database of authorized users or a data authentication server. Once authenticated, a user is usually subjected to an authorization process to determine whether the authenticated user entity should be permitted access and protected to use the system based on their privileges.

In this sub-step, the authentication process is required to indicate the users' groups and levels, namely: administrator, data analyst (experts), and decision-makers. Figure 6-4 gives an example of the data analyst or expert's privilege level. Data analysts can analyse and add the decision alternatives. Experts can add the quantitative data needed for analyzing the datasets.

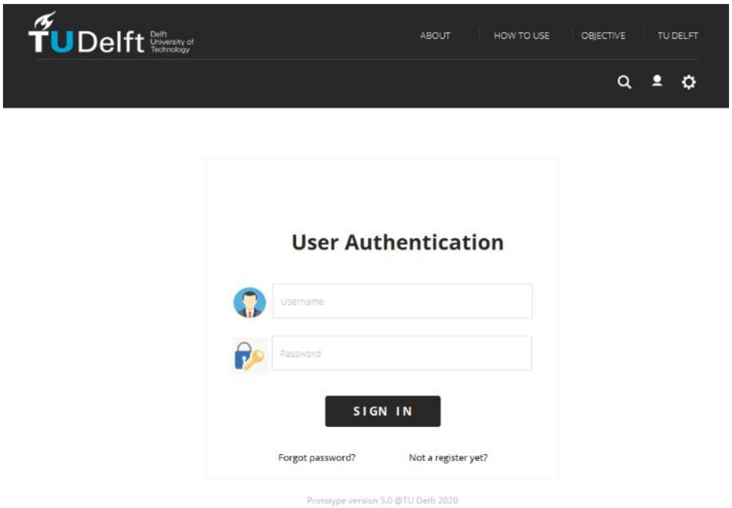


Figure 6-4 Sign-in process for the DSOD user

6.4.1.2 Message box

A message box is a dialogue box used that presents a message to the user. In this sub-step, the user can see the received mails from the DSOD's administrator. This message box aims to notify the users of the current status of the previous data analysis process. The DSOD will send the new instructions or tasks to the users. For example, the regular schedule to analyse the new dataset or reiterate and update the previous datasets. Therefore, this feature help users to be aware of the status and tasks in the DSOD prototype.

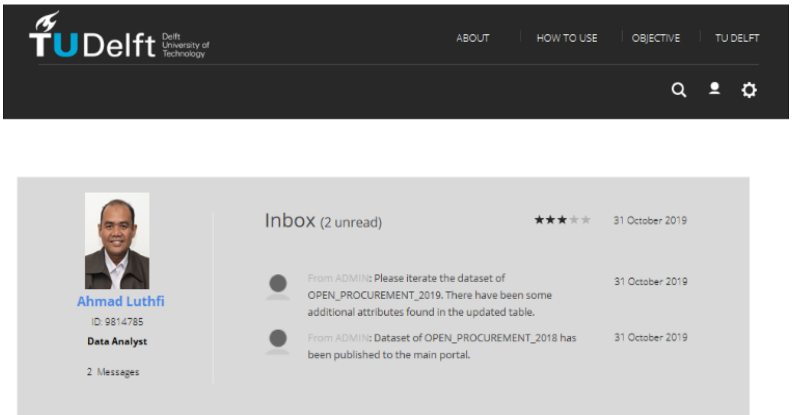


Figure 6-5 Message box for the user

6.4.1.3 Dataset selection

In this sub-step, data selection is defined as determining the appropriate data type and source, including the suitable preferred types of the file format to collect data. Data selection predates the actual practice of data collection. The main objective of data selection is to determine the appropriate data type, source, and instrument that allow investigators or analysts to analyse the selected dataset adequately.

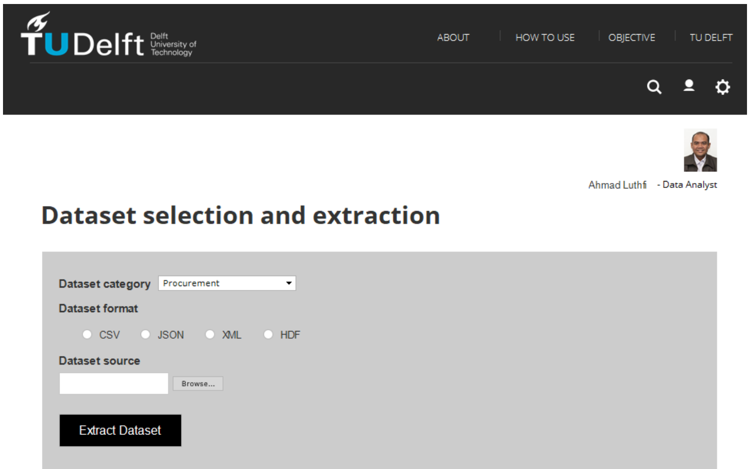


Figure 6-6 Dataset selection and extraction interface

Figure 6-6 shows that the tool selects the datasets from the data provider. The original dataset structure used in this case study is derived from government agencies. The decision-making support prototype will extract the selected dataset into a readable and machine structure in this process. The tool can select a data source from multiple database platforms like CSV, XML, JSON, etc., and ensure that the dataset's metadata is well structured.

6.4.1.4 Dataset extraction

In this sub-step, the DSOD provides the extraction process to retrieve and decompose the selected datasets. Data extraction is a process that involves the retrieval of data to process it further and store the data in a data-specific repository (Gour et al., 2010).

Data extraction is a process of retrieving datasets that are frequently unstructured to a structured and machine-readable format (Vassiliadis & Simitsis, 2009). Figure 6-7 presents the metadata of the dataset resulted from the extraction process. The extracting system is followed by dataset transformation to generate the datasets' metadata structure in this process. The metadata provided in this process consists of ten categories (see Table 6-1).

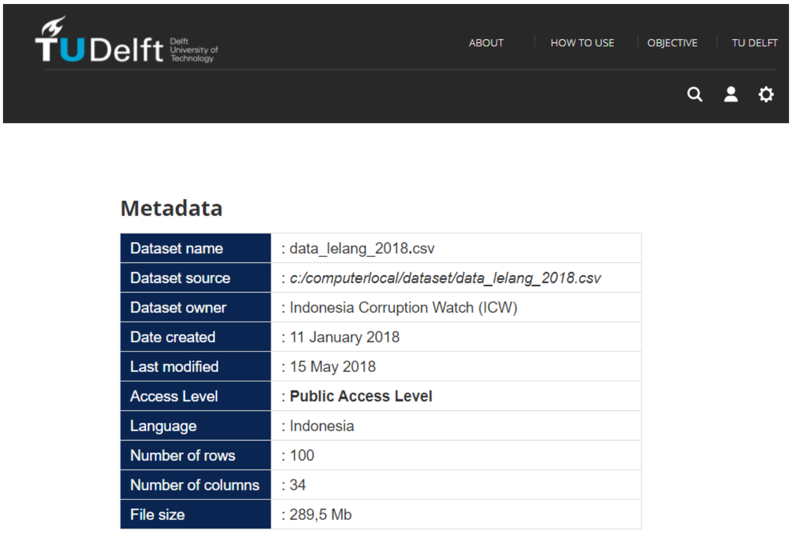


Figure 6-7 Dataset extraction and metadata interface

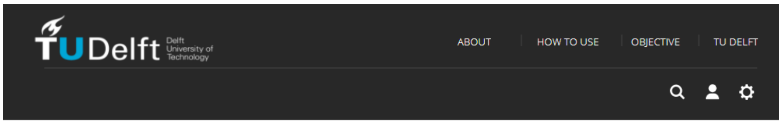
To categorise the metadata in the data extracting process, we used ten categories of Duval et al. (Duval, Hodgins, Sutton, & Weibel, 2002; Pomerantz, 2015) that helps to understand better the information about the dataset. The metadata categories are decomposed into dataset name, dataset source, dataset owner, date created, last modified, access level, language, number of rows, number of columns, and file size.

Table 6-1 Metadata categories and their functionality

	Metadata category	Description of functionality
1	Dataset name	Dataset name refers to the original name of the dataset file. The file includes the format used, such as CSV, JSON, XML, or DBF. This category's objective is to inform the users in terms of the name of the origin file. The DSOD allows the user to see what kinds of file types to analyse.
2	Dataset source	Data source refers to the original location of the dataset. The location source can be in the local computer machine or available at the distributed server systems. This category's objective is to ensure that the DSOD derives the file in the proper computer machine.
3	Dataset owner	Data owner refers to the legalization of the dataset. The DSOD will help to inform the owner of the dataset. This category's objective is to ensure that the dataset is legal and avoid property rights violations.
4	Date created	Date created refers to the original date and time of generated and elicited the dataset. The objective of this category is to inform the time of making the dataset to the users.
5	Last modified	Last modified refers to the latest time of modification or update of the dataset. This category aims to inform the timeline of the changes in the selected dataset.
6	Access level	Access level refers to the level status of the dataset. There are two access levels designed in the DSOD. First, the public access level means that the users can use and analyse the dataset publicly. Second, the private access level represents that the dataset only can be used by a limited number of users.
7	Language	Language refers to the metadata's language, including the attribute name and the dataset records' value. This category's objective is to inform the users which language setting is used in the selected dataset.
8	Number of rows	The number of rows refers to how many records containing in the dataset.
9	Number of columns	The number of rows refers to how many attributes or fields are containing in the dataset.
10	File size	The file size refers to the capacity of the selected dataset.

6.4.1.4 Dataset attributes

There are two types of attribute information provided in this step. First, attribute name indicates the original attribute names and what data is contained in each column of the selected dataset. Second, the data type of each attribute. The dataset attribute type can be shown in several formats, such as variable character (varchar), currency, and an integer (Noble, 2020). Varchar type is a set of character data of indeterminate length and refers to a data type that can accept letters and numbers. Currency is a monetary value given to data to identify its financial significance to an institution or organization. Integer, furthermore, is a data type that represents a specific range of mathematical integers and may or not may not be allowed to contain negative values. Figure 6-8 shows the dataset attribute information to the users.



Dataset Attribute

Attribute Name	Data Type
NAMA_SATKER	VARCHAR (25)
NAMA_PEMENANG	VARCHAR (25)
PAGU	CURRENCY
NPWP_PEMENANG	VARCHAR (15)
NAMA_KABUPATEN	VARCHAR (25)
KD_PEMENANG	VARCHAR (10)
NAMA_PAKET	VARCHAR (25)

Attribute Name	Data Type
NILAI_KONTRAK	CURRENCY
NAMA_AGENCY	VARCHAR (25)
HPS	CURRENCY
KODE_LPSE	VARCHAR (10)
NAMA_PANITIIA	VARCHAR (25)
SKOR_PEMENANG	INTEGER
SKOR_TOTAL	INTEGER
SKOR_TOTAL	INTEGER

Figure 6-8 Dataset attribute information

6.4.2 Analysing

The analysis step is a critical process in the DSOD, where the datasets will go through the evaluation process. In this step, the DSOD will support moving each data value from a selected table to the two broad categories of advantages and disadvantages. Thereafter, the dataset is evaluated using three methods (BbN, FMCDM, and DTA).

For example, in this step, the BbN method will assess all the selected attribute names by considering the probability and dependence of one attribute to another.

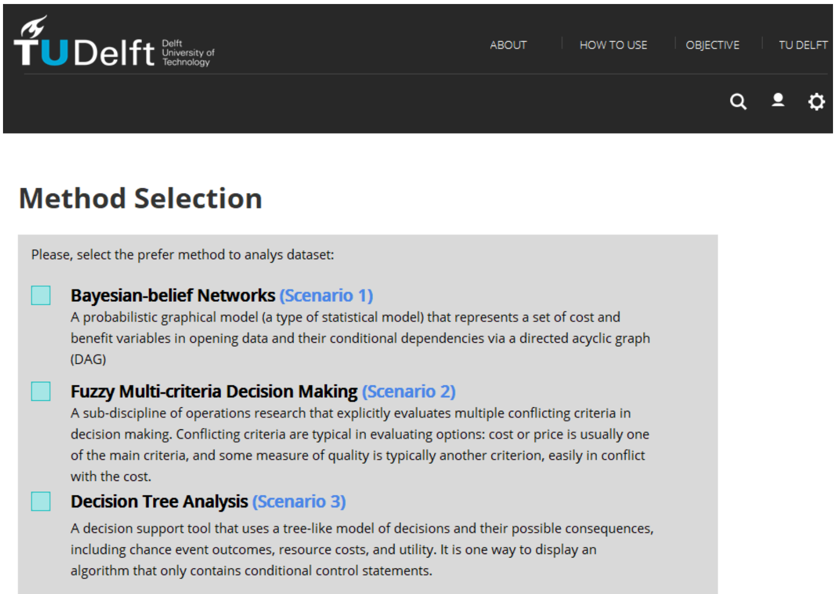


Figure 6-9 Methods selection interface

In this step, the DSOD provides three different selection methods (Bbn, FMCDM, and DTA). For instance, the prototype users, data analytics, decision-makers, researchers, or civil servants can select one of the methods or select the methods altogether to process the dataset simultaneously.

6.4.2.1 The use of Bayesian-belief Networks

In this method, there are five main steps to analyse the dataset. First, the DSOD will determine each attribute's advantages and disadvantages factors. In this step, the user can select the possible advantages and disadvantages factors based on their knowledge and expertise. Second, the DSOD will develop a causal network structure to show each attribute's cause and effect regarding the advantages and disadvantages factors. Third, the DSOD will formalize the structure to subjective prior beliefs from experts about the potential costs and benefits of opening data. The formulation to compute the probabilities of the cost and benefits factors of opening data. Fourth, the DSOD will quantify the prior possibility of defining and estimating the probability distributions for each benefit and cost factor. Fifth, the DSOD will interrogate the belief network to analyse the sensitivity level of the cost factors and present the high, moderate, and low probabilities of the potential investment. Figure 6-10 shows the analysis steps of BbN.

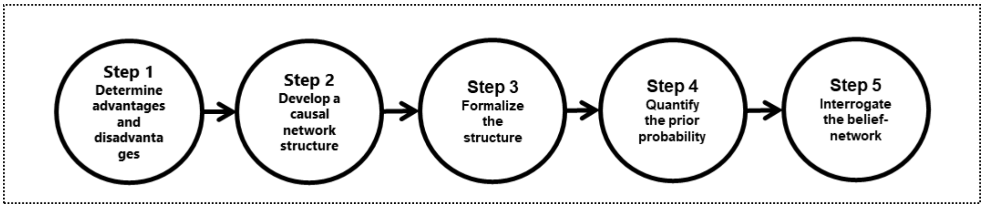


Figure 6-10 Analysis steps of the Bayesian-belief Networks

6.4.2.1.1 Determine the advantages and disadvantages

The main objective of this process is to determine the disadvantages and benefit factors. This step is presented in Figure 6-11. The tool asks the data analyst or expert to select single or multiple advantages and disadvantages category of the attribute. For example, the attribute name of *Nama_Satker* shown in Figure 6-11 is defined as multiple potential disadvantages, such as privacy infringement, data sensitivity, and personal identity. In this step, the DSOD prototype asks the data analyst or expert to choose a single or numerous disadvantage category of the attribute. This tool has several disadvantage categories: privacy infringement, data inaccuracy, data misinterpretation, data overlapping, data duplication, data sensitivity, data ownership, personal identity, incomplete data, and against the law.

At the same time, the DSOD prototype supports the expert and data analyst to choose a single or multiple benefits categories of the attribute. For example, the attribute name of *Nama_Pemenang* shown in Figure 6-11 has several potential advantages, such as increasing transparency, reputation, and accountability. This tool has several benefit categories derived from the developed taxonomy: increased transparency, citizen engagement, innovation, data reusability, data availability, reputation, better services, improved business processes, better understanding, data authenticity, and accountability.

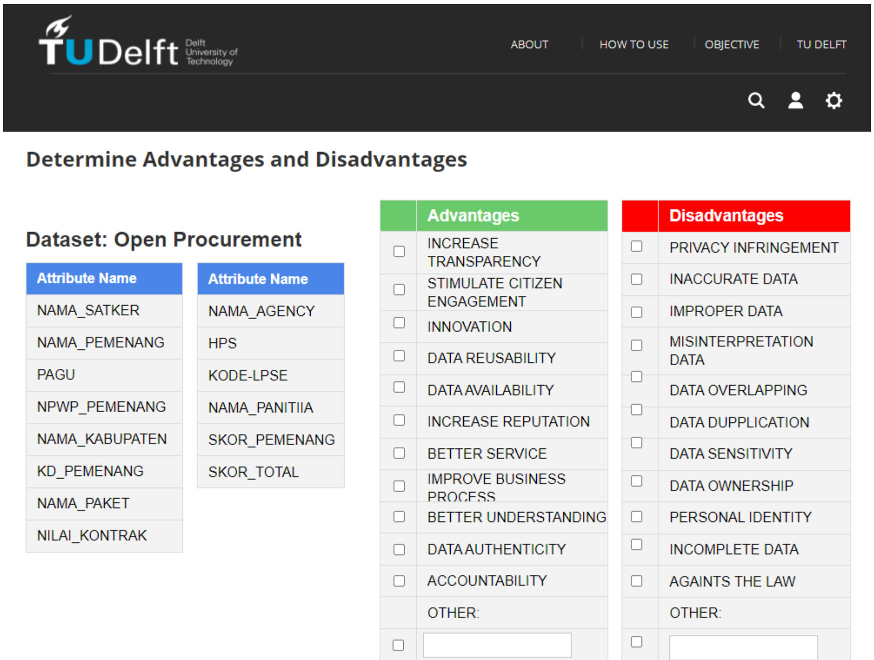


Figure 6-11 Determine the advantage and disadvantage categories

6.4.2.1.2 Develop a causal network structure

The Bayesian-belief Network structure is developed by identifying the causes and relationships between advantages and disadvantages in this step. We have identified three main disadvantage and benefit factors for the opening data: sensitivity, ownership, and improper data. Based on the cause-and-effect, a Bayesian-belief Network structure is created. This step can be processed in several iterations until the latest sub-node is identified and correlated.

Figure 6-12 illustrates the causalities of the disadvantage factors. There are three main disadvantage factors: data sensitivity, improper data, and data ownership (see Figure 6-12). Data sensitivity refers to releasing data that can include sensitive attributes. The users can analyse personal identity elements, like full name, date of birth, address, and phone number. Improper data refers to the releasing of data by the data providers or companies that are likely to drive a misinterpretation of the data. The causes factors of this category are: (a) insufficient domain expertise, (b) essential variables are omitted, (c) inappropriate data visualisation, and (d) error of attribute correlation. The effect of this disadvantage category is influencing the data quality and data incompleteness. Data ownership refers to the legal aspect and comprehensive control over a single piece of dataset elements. The inaccurate information about the data publishers' rightful owner of the datasets might ignore the acquisition and distribution policy of the datasets.

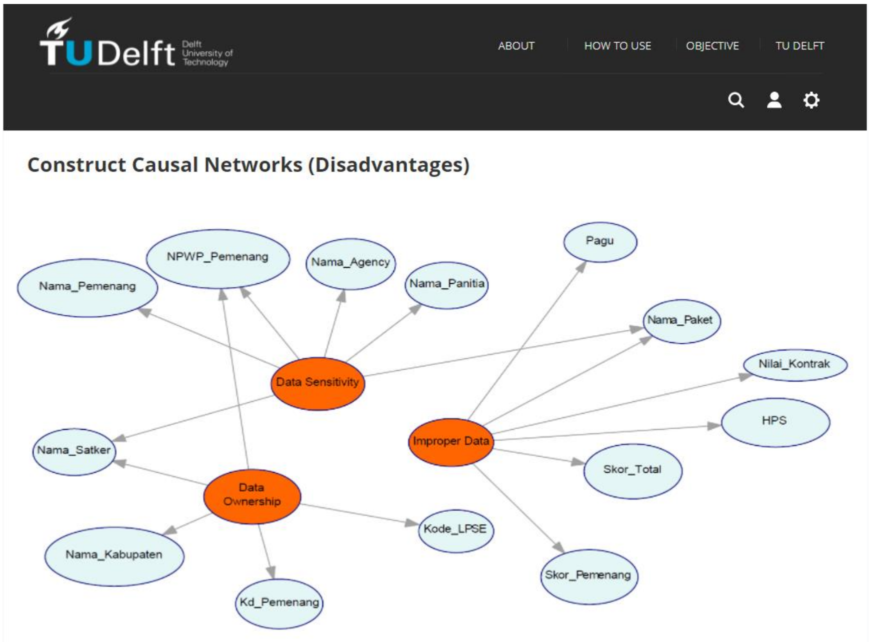


Figure 6-12 Construct the causalities of the disadvantage factors

Figure 6-13 illustrates the causalities of the advantage factors. There are three main advantage categories, namely transparency, reusability, and availability. Transparency refers to the sharing of the datasets, increasing the clarity of the government and individual performance. Society is being able to access relevant information through specific datasets. This situation can improve the decision-making process and save the investment of money both by the government and society. The reusability refers to the availability of the data in the government data portal is beneficial for the public to expose variability and enable experimentation. The public can re-use the data to generate new ideas or knowledge based on the experimental data. The availability refers to the process of ensuring that data is available to end-users without restriction. Providing high data availability can accelerate the opening of stored data.

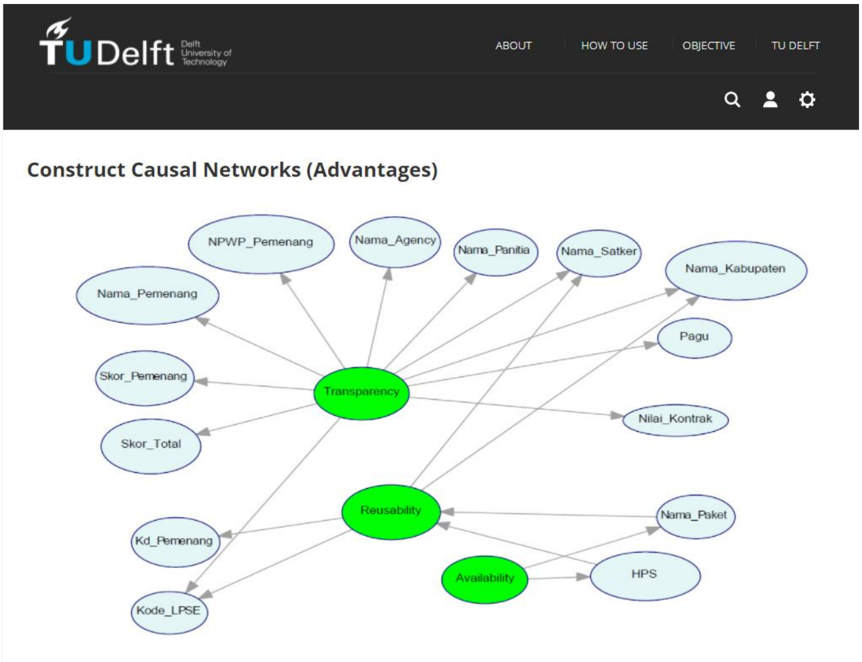


Figure 6-13 Construct the causalities of the benefit factors

6.4.2.1.3 Formulise the structure

The BbN can capture the uncertainty in the dependencies between the defined variables using conditional probabilities (Cárdenas et al., 2012). The probability factors in the BbN can compute the effect of the selected variable from the likelihood of a given cause element. The expert's judgment in this formulation refers to the experts' subjective prior beliefs about the potential costs and benefits of opening data. The formulation to compute the probabilities of the cost and benefits factors of opening data can be defined as follows:

P[effect]=[P[effect/cause].P[cause]]/P[cause/effect]

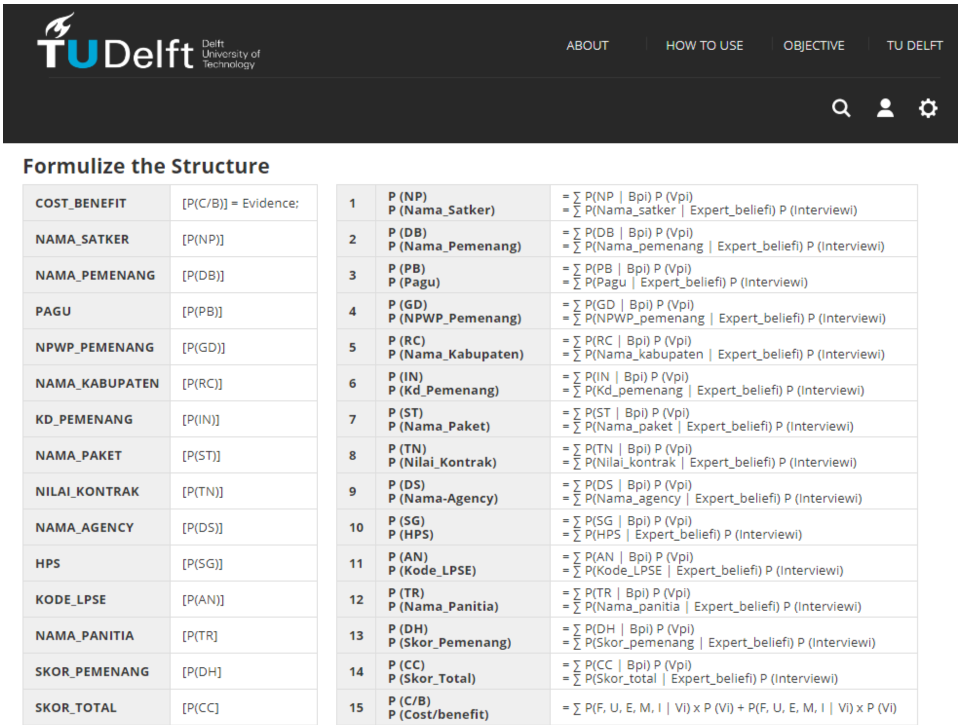
Where:

P[cause] = probability that the cause occurs,

P[effect] = probability that the effect occurs,

P[effect/cause] = conditional probability of the effect, given the cause,

P[cause/effect] = conditional probability of the cause, given the effect.



Formulize the Structure

COST_BENEFIT	[P(C/B)] = Evidence;
NAMA_SATKER	[P(NP)]
NAMA_PEMENANG	[P(DB)]
PAGU	[P(PB)]
NPWP_PEMENANG	[P(GD)]
NAMA_KABUPATEN	[P(RC)]
KD_PEMENANG	[P(IN)]
NAMA_PAKET	[P(ST)]
NILAI_KONTRAK	[P(TN)]
NAMA_AGENCY	[P(DS)]
HPS	[P(SG)]
KODE_LPSE	[P(AN)]
NAMA_PANITIA	[P(TR]
SKOR_PEMENANG	[P(DH]
SKOR_TOTAL	[P(CC]

1	P (NP) P (Nama_Satker)	= ∑ P(NP | Bpi) P (Vpi) = ∑ P(Nama_satker | Expert_beliefi) P (Interviewi)
2	P (DB) P (Nama_Pemenang)	= ∑ P(DB | Bpi) P (Vpi) = ∑ P(Nama_pemenang | Expert_beliefi) P (Interviewi)
3	P (PB) P (Pagu)	= ∑ P(PB | Bpi) P (Vpi) = ∑ P(Pagu | Expert_beliefi) P (Interviewi)
4	P (GD) P (NPWP_Pemenang)	= ∑ P(GD | Bpi) P (Vpi) = ∑ P(NPWP_pemenang | Expert_beliefi) P (Interviewi)
5	P (RC) P (Nama_Kabupaten)	= ∑ P(RC | Bpi) P (Vpi) = ∑ P(Nama_kabupaten | Expert_beliefi) P (Interviewi)
6	P (IN) P (Kd_Pemenang)	= ∑ P(IN | Bpi) P (Vpi) = ∑ P(Kd_pemenang | Expert_beliefi) P (Interviewi)
7	P (ST) P (Nama_Paket)	= ∑ P(ST | Bpi) P (Vpi) = ∑ P(Nama_paket | Expert_beliefi) P (Interviewi)
8	P (TN) P (Nilai_Kontrak)	= ∑ P(TN | Bpi) P (Vpi) = ∑ P(Nilai_kontrak | Expert_beliefi) P (Interviewi)
9	P (DS) P (Nama-Agency)	= ∑ P(DS | Bpi) P (Vpi) = ∑ P(Nama_agency | Expert_beliefi) P (Interviewi)
10	P (SG) P (HPS)	= ∑ P(SG | Bpi) P (Vpi) = ∑ P(HPS | Expert_beliefi) P (Interviewi)
11	P (AN) P (Kode_LPSE)	= ∑ P(AN | Bpi) P (Vpi) = ∑ P(Kode_LPSE | Expert_beliefi) P (Interviewi)
12	P (TR) P (Nama_Panitia)	= ∑ P(TR | Bpi) P (Vpi) = ∑ P(Nama_panitia | Expert_beliefi) P (Interviewi)
13	P (DH) P (Skor_Pemenang)	= ∑ P(DH | Bpi) P (Vpi) = ∑ P(Skor_pemenang | Expert_beliefi) P (Interviewi)
14	P (CC) P (Skor_Total)	= ∑ P(CC | Bpi) P (Vpi) = ∑ P(Skor_total | Expert_beliefi) P (Interviewi)
15	P (C/B) P (Cost/benefit)	= ∑ P(F, U, E, M, I | Vi) x P (Vi) + P(F, U, E, M, I | Vi) x P (Vi)

Figure 6-14 Formulise the structure

6.4.2.1.4 Quantify the prior probability

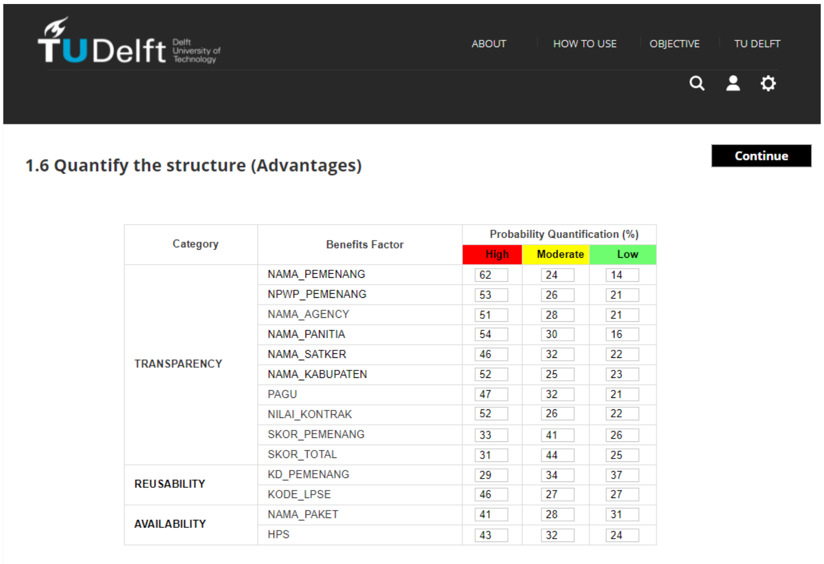
The objective of calculating posterior probabilities is defining and estimating the probability distributions for each advantage and disadvantage factors. There are two main procedures to quantify the posterior probability factors. First, select the experts' team based on their formal education, functional knowledge, and practical insight. Second, quantify the advantages and disadvantages factors by the experts' judgment.

a. Experts' domain and expertise

The experts' selection is based on their formal education, functional knowledge, and practical insight. The experts need to accommodate various specializations that partially overlap to confirm the completeness of the data or information available (Herland et al., 2016; Teicher, 2015).

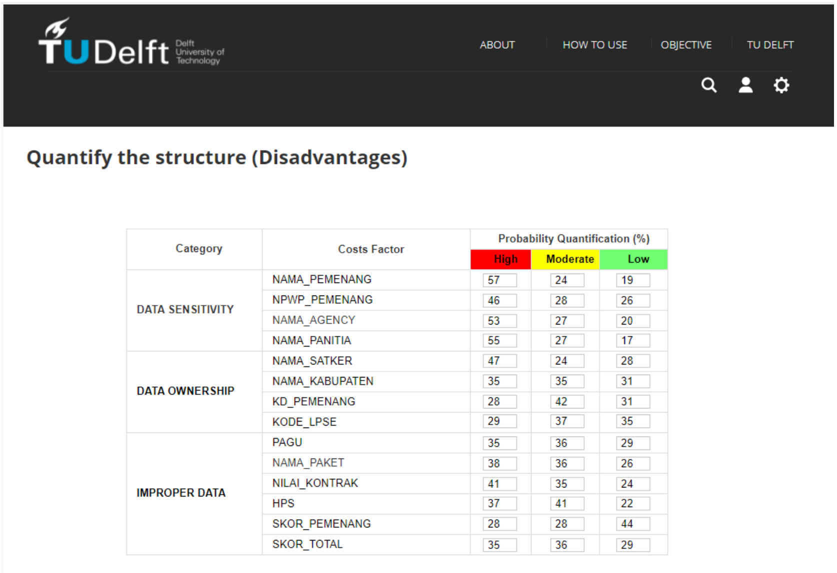
b. Experts' Judgment Quantification

Expert judgment quantification required numerical data representing the event frequencies, causal relationships, and conditional probabilities for the benefits and disadvantages of opening data. Figures 6-15 and 6-16 show the quantification process of the advantage and disadvantage factors. In this step, the selected experts quantify the posterior probabilities to estimate the probability distribution level, namely High, Moderate, and Low of the advantage and disadvantage factors.



Category	Benefits Factor	Probability Quantification (%)		
		High	Moderate	Low
TRANSPARENCY	NAMA_PEMENANG	62	24	14
	NPWP_PEMENANG	53	26	21
	NAMA_AGENCY	51	28	21
	NAMA_PANITIA	54	30	16
	NAMA_SATKER	46	32	22
	NAMA_KABUPATEN	52	25	23
	PAGU	47	32	21
	NILAI_KONTRAK	52	26	22
	SKOR_PEMENANG	33	41	26
	SKOR_TOTAL	31	44	25
REUSABILITY	KD_PEMENANG	29	34	37
	KODE_LPSE	46	27	27
AVAILABILITY	NAMA_PAKET	41	28	31
	HPS	43	32	24

Figure 6-15 Quantify the advantages



Quantify the structure (Disadvantages)

Category	Costs Factor	Probability Quantification (%)		
		High	Moderate	Low
DATA SENSITIVITY	NAMA_PEMENANG	57	24	19
	NPWP_PEMENANG	46	28	26
	NAMA_AGENCY	53	27	20
	NAMA_PANITIA	55	27	17
DATA OWNERSHIP	NAMA_SATKER	47	24	28
	NAMA_KABUPATEN	35	35	31
	KD_PEMENANG	28	42	31
	KODE_LPSE	29	37	35
IMPROPER DATA	PAGU	35	36	29
	NAMA_PAKET	38	36	26
	NILAI_KONTRAK	41	35	24
	HPS	37	41	22
	SKOR_PEMENANG	28	28	44
	SKOR_TOTAL	35	36	29

Figure 6-16 Quantify the disadvantages

6.4.2.1.5 Interrogate the belief-network

In this step, we interrogated the resulting model by distributing the probabilities for each node of the disadvantage factors and their sub-factors. This step interrogates the cost factors' sensitivity level and presents the high, moderate, and low potential investment probabilities. Figure 6-17 shows the causal relationship between factors influencing the disadvantaged category and its sub-node probability. Based on the quantification process in Figure 6-16, the highest possible disadvantage is on the data sensitivity factor (58%). The data sensitivity is the most influenced factor affected by nama_pemenang, NPWP_pemenang, Nama_agency, and Nama_panitia. The status degree of the disadvantage, in this case, is moderate (34%) for improper data, which means of releasing the selected dataset tends to have a moderate adverse effect on the data providers, especially in the issues of data ownership and improper data.

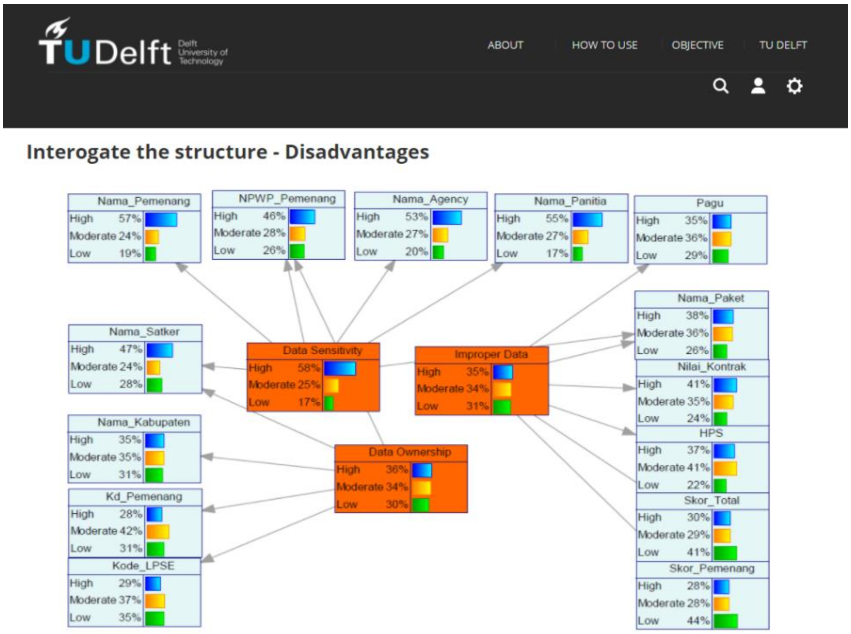


Figure 6-17 Interrogate the disadvantages

Figure 6-17 reveals the causal relationship between factors influencing the disadvantages of opening data derived from the quantification process in Figure 6-16. The figure shows that the highest possible benefit is in the transparency factor (53%). Further, the citizen engagement factor is influenced by the communication exchange and interactive design. The advantage status level is high (41%). This means that releasing the selected dataset might have highly relevant advantages.

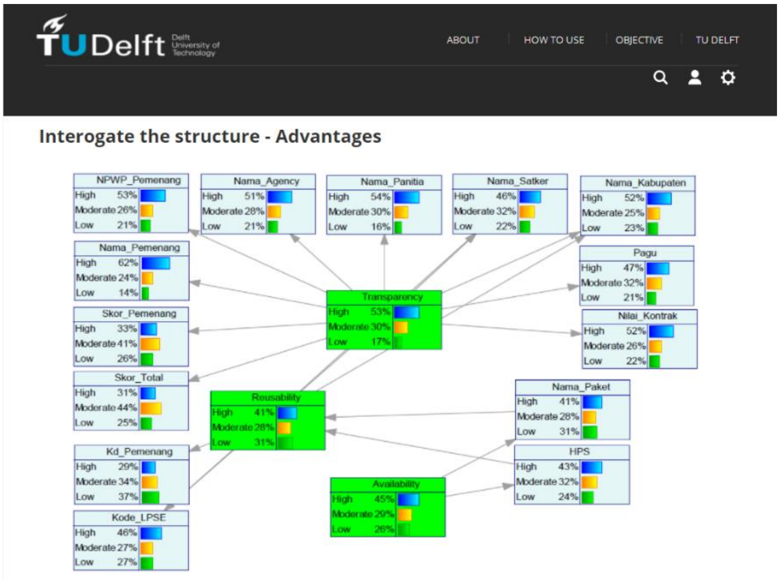


Figure 6-18 Interrogate the benefit

6.4.2.2 The use of Fuzzy Multi-Criteria Decision Making

In this method, there are four main steps to analyse the dataset. First, the DSOD will determine the possible criteria and sub-criteria of each attribute's advantages and disadvantages factors. In this step, the user can select the possible disadvantages and benefits criteria based on their knowledge and expertise. Second, the DSOD will construct a relationship diagram to show the relationship between criteria and sub-criteria of the advantages and disadvantages factors. Third, the DSOD will construct the hierarchical diagram to show the criteria, sub-criteria, and decision alternatives. Fourth, the DSOD will define decision alternatives for each criterion and sub-criteria. Figure 6-19 shows the analysis steps of the FMCDM.

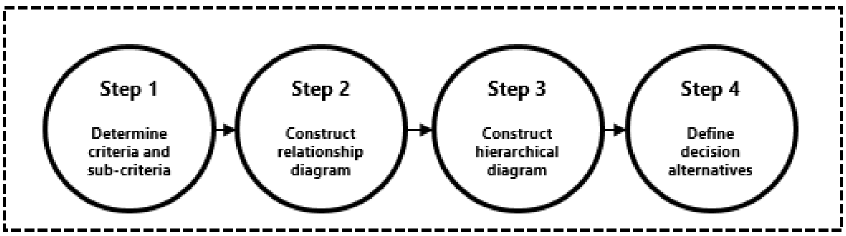


Figure 6-19 Analysis steps of the FMCDM

6.4.2.2.1 Determine criteria and sub-criteria

Figure 6-20 represents the hierarchy of the four criteria, eight sub-criteria, and four alternatives. There are three main cost criteria, namely data sensitivity, data ownership, and data overlapping. Data sensitivity refers to releasing data that can include sensitive attributes. Personal identity elements, like full name, date of birth, address, and phone number, are possible to be analysed. Data ownership refers to the rights and comprehensive control over a single piece of dataset elements. The inaccurate information about the data publishers' rightful owner of the datasets is needed to know if the data might be opened. Data overlapping refers to the datasets that might contain overlapping collections of data. More datasets on various government portals might include data on a similar theme or subject.

Furthermore, on the benefit side, three main criteria might reap from the opening of the dataset. First, data availability refers to the process of ensuring that data is available to end-users without restriction. Providing high data availability can accelerate stored data to be accessible to anyone and valid in the real-time process. Second, data trustworthiness refers to how the public sector can better use open data to inform citizens about their actions. By publication of open data, a public-sector body can present itself as an open and transparent institution. Third, data reusability refers to if the data available in the government data portal is beneficial for the public to expose variability and enable experimentation. The public can re-use the data to generate new ideas or knowledge based on the experimental data.

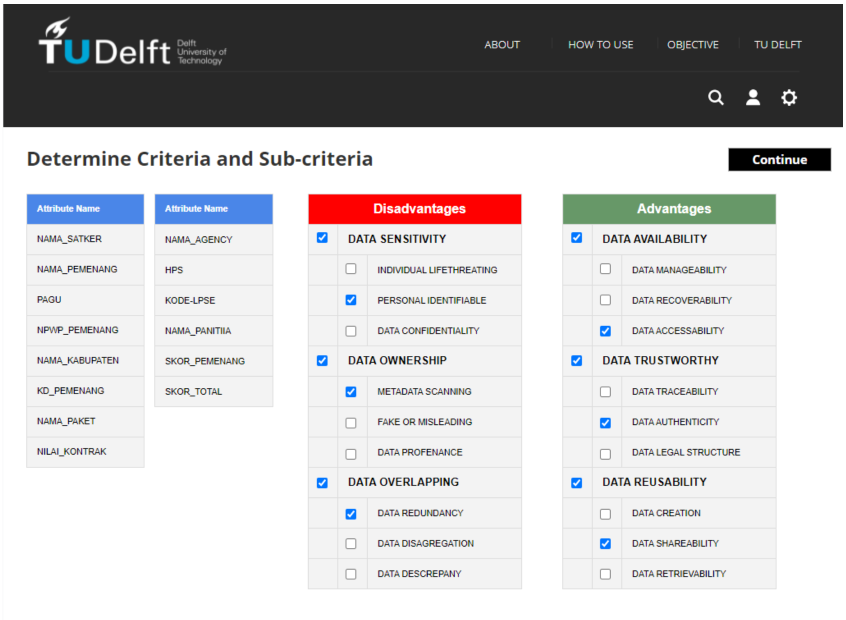


Figure 6-20 Determine criteria and sub-criteria

6.4.2.2.2 Construct relationship diagram

The DSOD will construct the relationship diagram to develop a hierarchical relationship between the factors in this step. The data sensitivity (C1) is composed of two sub-criteria: individual life-threatening (C1.1) and data identifiable (C1.2). Individual life-threatening (C1.1) can be defined as a potential disadvantage to an individual or personal life because of the possibility of recognising the dataset's sensitive value. Data identifiable (C1.2) is specified as the possible leak of the personal, organisational, business, or even government data identifiable, e.g., by combining some field attributes.

The second criterion is data ownership (C2), which consists of two sub-criteria, namely metadata scanning (C2.1) and fake or misleading (C2.2). Metadata scanning (C2.1) can be represented to figure out the property and structure of the dataset. Fake or misleading (C2.2) is defined by a user to potentially change and modify the dataset and affect an unreliable and wrong decision. Data availability (C3) is the third criterion, and it has two sub-criteria, namely, data manageability (C3.1) and data recoverability (C3.2). Data manageability (C3.1) is specified as the chance to manage the dataset's availability and accessibility. Data recoverability (C3.2) is indicated by delivering a dataset, and it can have a highly positive impact on recovering the availability of the data. The fourth criterion is data trustworthiness (C4), which consists of two sub-criteria, e.g., data traceability (C4.1) and data authenticity (C4.2). Data traceability (C4.1) can make the possibility to trace the source of the dataset. Data authenticity (C4.2) is defined as potentially affected to recognise the authentication of the data.

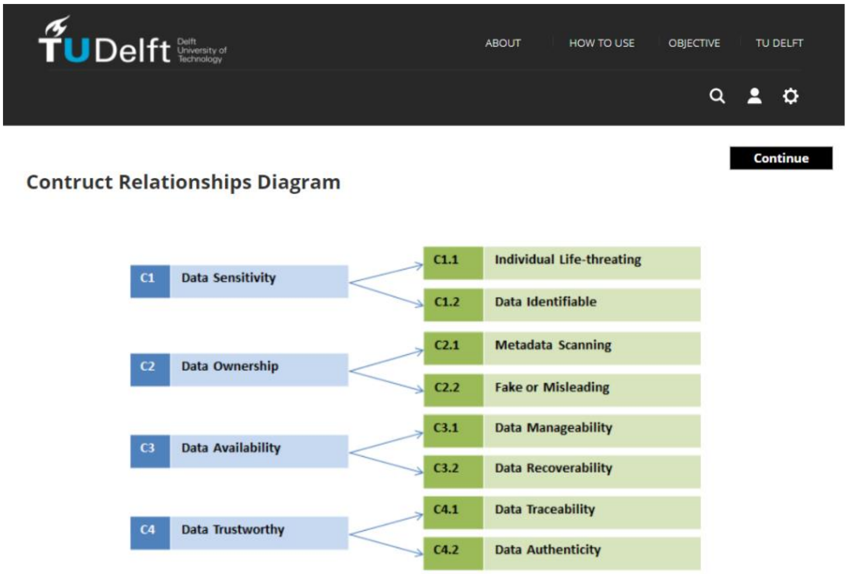


Figure 6-21 Construct relationship diagram

6.4.2.2.3 Construct a hierarchical diagram

Figure 6-22 represents the hierarchy of the four criteria, eight sub-criteria, and four alternatives. The four criteria, C1, C2, C3, and C4, define data sensitivity, ownership, data availability, and trustworthiness. The data sensitivity (C1) composes of two sub-criteria: individual life-threatening (C1.1) and data identifiable (C1.2), as discussed earlier.

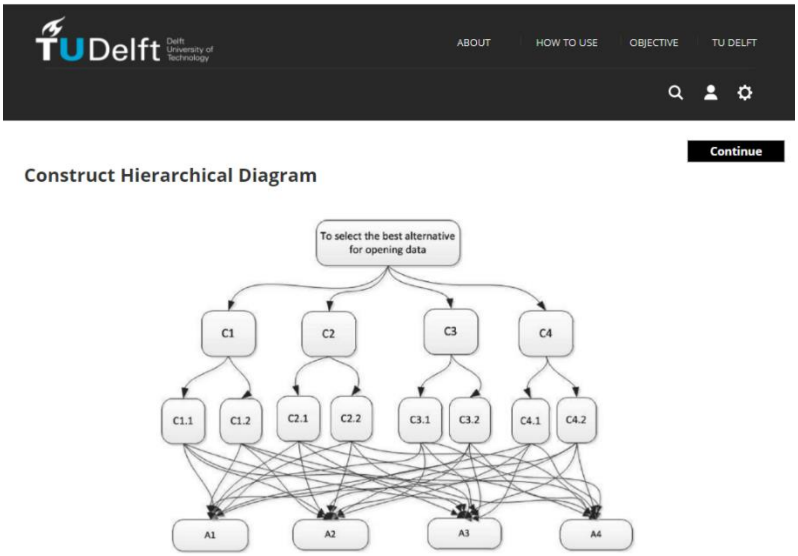


Figure 6-22 Construct a hierarchical diagram

6.4.2.2.4 Define decision alternatives

The following four alternatives of opening data in this paper are: opening the dataset (A1), maintaining a dataset suppression (A2), providing limited access (A3), or keeping the dataset closed (A4). First, “open the dataset” refers to the publishing of the dataset. In this situation, the opening potentially presents a low disadvantage to an individual or organisation, or the potential benefits of the dataset substantially outweigh the potential disadvantages. Second, “maintaining suppression” means removing a data field and/or an individual record into particular groups or generating unique characteristics to avoid personal identity. In this alternative, data might create significant disadvantages and should not be opened in the actual form, as the potential benefits do not outweigh the disadvantages. Third, the alternative “limited access” defines that only a certain group will be given access to the data. The level of openness is limited. Often those who will gain access have to sign a document that outlines the rules of access. This is because releasing the dataset will create a moderate disadvantage, or the potential benefits of the dataset do not outweigh the potential privacy disadvantages. Fourth, the alternative “keeping the dataset closed” means publishing the dataset generates a very high disadvantage and significantly outweighs the potential advantages.

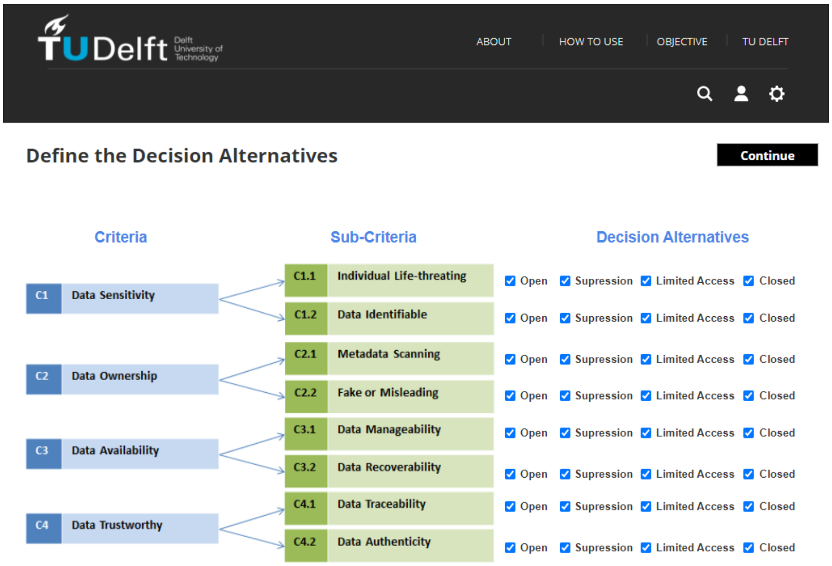


Figure 6-23 Define decision alternatives

6.4.2.3 The use of Decision Tree Analysis

In this method, there are two main steps to analyse the dataset. First, the DSOD will determine the advantages and disadvantages categories. In this step, the user can select the possible advantages and disadvantages criteria based on their knowledge and expertise. Second, the DSOD will construct a decision structure to show the tree structure of alternative decisions and their disadvantages and benefits categories. Figure 6-24 shows the analysis steps of the DTA.

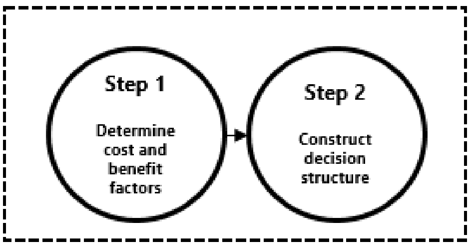


Figure 6-24 Analysis steps of the Decision Tree Analysis

6.4.2.3.1 Determine advantages and disadvantages

The first step in analysing step of the DTA is determining the cost and benefit factors. In this step, the DSOD will select all the possible cost and benefits factors by default. However, the user or analyst can choose the preferred cost and benefit factors based on their knowledge and expertise. In the open decision, categories such as the cost for data collection, data visualisation, and treating sensitive data are the three main costs for opening datasets. At the same time, opening data can also contribute to the benefits, such as increasing community engagement and improving the data providers' accountability.

In the limited access decision, factors like the cost of suppressing data, maintaining data, and protecting personal identity are the main issues in making this decision. Meanwhile, a "limited access" decision can give several benefits, such as providing the data's confidentiality, ensuring the users' authentication, and delivering a new knowledge of a better understanding of the open data users.

Finally, the main problems are in the "closed" decision, such as the cost for preserving data, providing a storage technology, and updating the incomplete dataset. Simultaneously, the closed decision can contribute to the advantages, such as protecting the privacy of individuals.

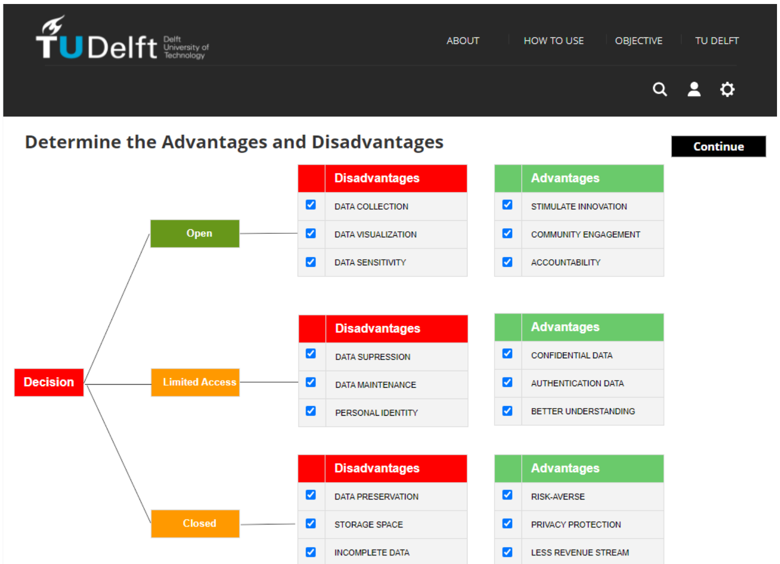


Figure 6-25 Determine advantages and disadvantages category

6.4.2.3.2 Construct decision structure

The decision-making process in opening data can be time-consuming and might require many resources. The decision-makers need to simplify the complex and strategic challenges to better understand each possible outcome's consequences. Therefore, the DTA can construct a model and structure the decision alternatives, whether the data should be released or closed. Figure 6-26 illustrates the decision alternatives and various possible paths in deciding the complex problems of opening data. We define three primary decision nodes, namely "open", "limited access", and "closed". The first "open" decision refers to the governments releasing their data to the public with fewer or no restrictions. Second, "limited access" indicates that the level of openness is restricted to a specific group of users. Third, a "closed" decision refers to the government should keep the data exclusively.

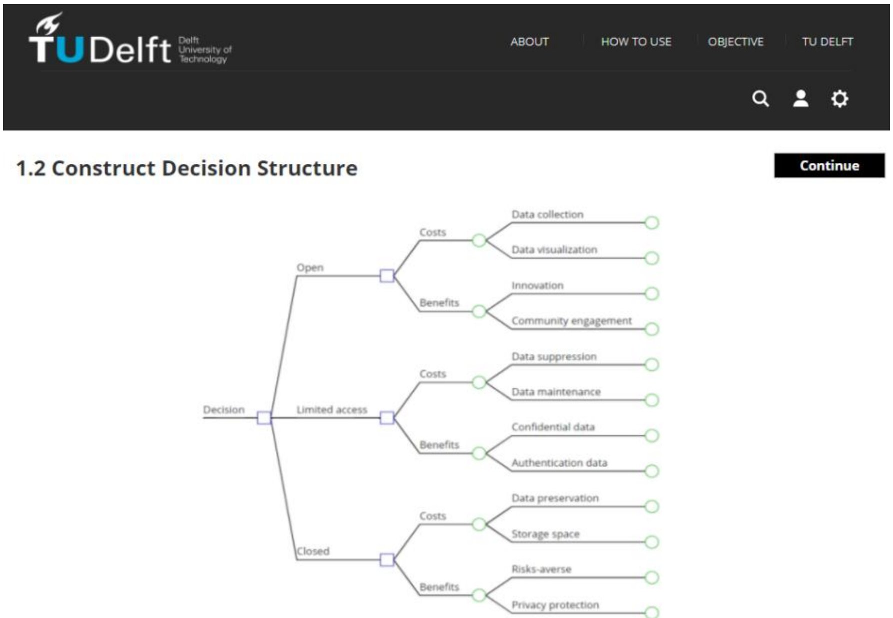


Figure 6-26 Construct decision structure

6.4.3 Weighing

This step's main objective is to define the selected dataset's qualitative and quantitative advantages and disadvantages. In this study, we use qualitative and quantitative approaches to facilitate and help the experts in case of less knowledge to define and categorise the benefits and disadvantages, including the risks and costs category. This system's advantage is to provide iterative process conditions to ensure that the benefits level is higher than the disadvantage at hand. This step also provides a method to define the possibility of occurrence and impact of the predictable benefits. There are five categories in the qualitative value to define the benefit and cost level, whereas the quantitative value is determined using scale numbers from 0 to 100.

Table 6-2 shows the qualitative and quantitative consequences for the benefits of the opening dataset. This research uses five qualitative and quantitative categories to represent the consequence levels and make it convenient for experts to weigh their judgments. These five qualitative value categories could impact the benefits of personal, society, governments, academia, community, business enablers, and other open data stakeholders. First, a very high benefit category refers to the dataset with multiple compelling and essential utilities. At this level, the advantage of opening data is almost certain to occur to the open data stakeholders. Second, a high benefit refers to the dataset that will likely have a compelling and essential utility. At this level, the advantage of opening data is highly likely to occur. Third, a moderate benefit refers to the dataset that will likely have a clear utility, which means that the advantage of opening data is somewhat likely to occur to the open data stakeholders. Fourth, a low benefit refers to the dataset that will likely have a limited utility. At this level, the advantage of opening data is unlikely to occur to the open data stakeholders. Fifth, a very low benefit represents the dataset will likely have little utility. In this category, the advantage of opening data is doubtful to occur to the open data stakeholders.

Table 6-2 The consequence of advantage levels to open data

Qualitative value (benefit consequence)	Quantitative value (benefit consequence)	Potential occur of opening data	Description
Very high benefit	81 – 100	The advantage of opening data is almost certain to occur.	The dataset will likely have multiple compelling and essential utilities for open government data stakeholders.
High benefit	61 – 80	The advantage of opening data is very likely to occur.	The dataset will likely have a compelling and essential utility for open government data stakeholders
Moderate benefit	41 – 60	The advantage of opening data is somewhat likely to occur.	The dataset will likely have a clear utility for open government data stakeholders.
Low benefit	21 – 40	The advantage of opening data is unlikely to occur.	The dataset will likely have a limited utility for open government data stakeholders
Very low benefit	0 – 20	The advantage of opening data is very unlikely to occur.	The dataset will likely have negligible utility for open government data stakeholders

Table 6-3 presents the possibility of occurrence and the impact of predictable benefits. The use of this table is to compare the likelihood of occurrence and the impact of the foreseeable benefits of opening data. For example, if the possible event is high and the impact of the predictable benefit is moderate, then the foreseeable benefit will be moderate.

Table 6-3 Possible occurrences and the impact of predictable benefits

Possibility of occurrence	Impact of predictable benefits				
	Very high impact	**High impact**	**Moderate impact**	**Low impact**	**Very low impact**
Very high possibility	Very high benefit	Very high benefit	High benefit	Moderate benefit	Low benefit
High possibility	Very high benefit	High benefit	Moderate benefit	Moderate benefit	Low benefit
Moderate possibility	High benefit	Moderate benefit	Moderate benefit	Low benefit	Low benefit
Low possibility	Moderate benefit	Moderate benefit	Low benefit	Low benefit	Very Low benefit
Very low possibility	Low benefit	Low benefit	Low benefit	Very Low benefit	Very Low benefit

Table 6-4 indicates the qualitative and quantitative values for the consequence of the opening dataset disadvantages. Five categories of the qualitative value could impact the various potential disadvantages to the personal, society, governments, academicians, community, business enablers, and other open data stakeholders. First, the very high-risk category refers to the dataset that will likely have multiple severe adverse effects. At this level, the disadvantage of opening data is almost certain to occur to the open data stakeholders. Second, a high risk refers to the dataset will likely have a severe adverse effect. At this level, the disadvantage of opening data is highly likely to occur. Third, a moderate disadvantage refers to the dataset will likely have a moderate adverse effect, which means that the disadvantages of opening data are somewhat likely to occur. Fourth, a low benefit refers to the dataset will likely have a limited adverse impact. At this level, the disadvantage is that opening data is unlikely to occur to the open data stakeholders. Fifth, a very low risk represents the dataset will likely have an insignificant adverse impact. In this category, the disadvantages of opening data are improbable to occur to the open data stakeholders.

Table 6-4 The consequence of disadvantages levels in the opening dataset

Qualitative value (disadvantage consequence)	Quantitative value (risk consequence)	Potential occurs of releasing data	Description
Very high disadvantage	81 – 100	The disadvantage is almost certain to occur.	The dataset will likely have multiple severe adverse effects for open data stakeholders.
High disadvantage	61 – 80	The disadvantage is very likely to occur.	The dataset will likely have a severe adverse effect on open data stakeholders.
Moderate disadvantage	41 – 60	The disadvantage is somewhat likely to occur.	The dataset will likely have a moderate adverse effect on open data stakeholders.
Low disadvantage	21 – 40	The disadvantage is unlikely to occur.	The dataset will likely have a limited adverse impact on open data stakeholders.
Very low disadvantage	0 – 20	The risk is very unlikely to occur.	The dataset will likely have an insignificant adverse impact on open data stakeholders.

Table 6-5 presents the possibility of occurrence and the impact of predictable disadvantages. The use of this table is to compare the likelihood of occurrence and the impact of the foreseeable disadvantages of opening data. For example, suppose the possible occurrence is high, and the predictable benefit's impact is moderate. In that case, the foreseeable disadvantages will be moderate according to the standard classification we included in the DSOD. Indeed, it is possible to modify the category by experts. The opening data's utilities show in Table 6-2 and Table 6-4 can be for personal uses, society benefits, government needs, academic purposes, community intentions, business projects, and other open data stakeholders.

Table 6-5 Possible occurrences and the impact of predictable disadvantages

Possibility of occurrence	Impact of predictable disadvantages				
	Very high impact	High impact	Moderate impact	Low impact	Very low impact
Very high possibility	Very high disadvantage	Very high disadvantage	High disadvantage	Moderate disadvantage	Low disadvantage
High possibility	Very high disadvantage	High disadvantage	Moderate disadvantage	Moderate disadvantage	Low disadvantage
Moderate possibility	High disadvantage	Moderate disadvantage	Moderate disadvantage	Low disadvantage	Low disadvantage
Low possibility	Moderate disadvantage	Moderate disadvantage	Low disadvantage	Low disadvantage	Very Low disadvantage
Very low possibility	Low disadvantage	Low disadvantage	Low disadvantage	Very Low disadvantage	Very Low disadvantage

6.4.3.1 The use of BbN in the prototype

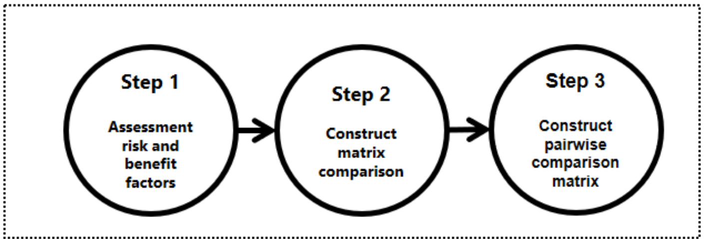


Figure 6-27 Weighing steps of the BbN

6.4.3.1.1 Assessment of advantages and disadvantages

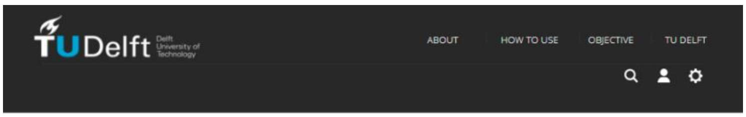
The estimated disadvantage and benefit will be analysed after quantifying and deriving the cost and benefit using the BbN approach. This process's main objective is to prepare for the decision to be made. The four steps of weighing the costs-benefits process are designed for the experts to give a vetting status structure (Altman, Wood, O'Brien, Vadhan, & Gasser, 2015).

To start, DSOD needs to evaluate the information in the dataset by combining this with the estimated advantages and disadvantages. Second, DSOD should show the potential benefits and users of the dataset against the likelihood of their evidence.

Third, experts evaluate the potential costs by analyzing the disadvantages of the dataset's adverse effects against the likelihood of their evidence. Last, the DSOD will process the weigh mechanism by integrating the overall quantified result from the previous step (Step 2).

Figure 6-28 presents the assessment result of the advantages and disadvantages factors. These values are derived from the analysing steps (step 4 – quantify the prior probability).

In this step, the expert quantified numerical data representing the event frequencies, causal relationship, and conditional probabilities in terms of disadvantages and advantages for opening data. For example, in the data sensitivity, the attribute of Nama_Pemenang was quantified as 62% to depict possible high disadvantage, 22% to potential moderate disadvantage, and 16% to possible low disadvantage. Therefore, Nama_Pemenang attribute tends to be a high potential disadvantage (62%). At the same time, the attribute of Nama_Pemenang can also reap advantages from the transparency side. The attribute of Nama_Pemenang quantified as 68% to possible high benefit, 22% to potential moderate advantage, and 10% to possible low benefit. Thus, Nama_Pemenang attribute tends to likely high benefit equivalent to 68%. From this result, we can see that the Nama_Pemenang attribute has a potentially high disadvantage and high benefit at the same time.



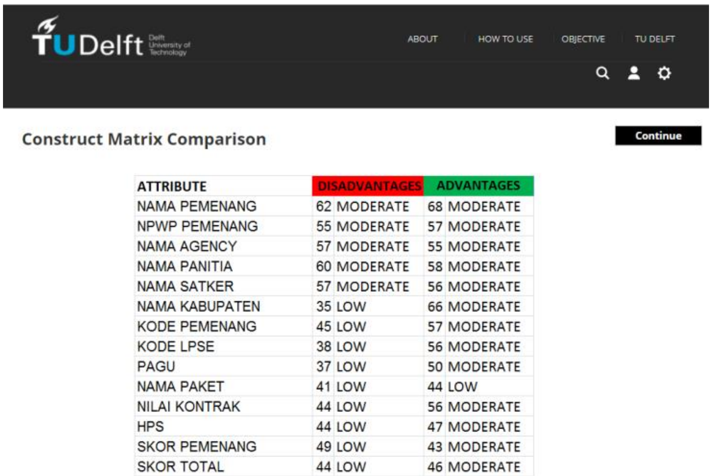
COST FACTORS	ATTRIBUTE	HIGH	MODERATE	LOW
DATA SENSITIVITY	NAMA PEMENANG	62	22	16
	NPWP PEMENANG	55	25	20
	NAMA AGENCY	57	26	17
	NAMA PANITIA	60	26	14
DATA OWNERSHIP	NAMA SATKER	57	18	25
	NAMA KABUPATEN	35	35	30
	KODE PEMENANG	25	45	30
	KODE LPSE	27	38	35
IMPROPER DATA	PAGU	36	37	27
	NAMA PAKET	41	38	21
	NILAI KONTRAK	44	36	20
	HPS	39	44	17
	SKOR PEMENANG	25	26	49
	SKOR TOTAL	29	27	44

BENEFIT FACTORS	ATTRIBUTE	HIGH	MODERATE	LOW
TRANSPARENCY	NAMA PEMENANG	68	22	10
	NPWP PEMENANG	57	25	18
	NAMA AGENCY	55	27	18
	NAMA PANITIA	58	29	13
	NAMA SATKER	56	31	13
	NAMA KABUPATEN	66	19	15
	PAGU	50	32	18
	NILAI KONTRAK	56	24	20
	SKOR PEMENANG	33	43	24
	SKOR TOTAL	31	46	23
REUSABILITY	KD PEMENANG	27	34	39
	KODE LPSE	56	22	22
AVAILABILITY	NAMA PAKET	44	26	30
	HPS	47	32	21

Figure 6-28 Assessment disadvantage and advantage factors

6.4.3.1.2 Construct comparison matrix

In this step, the DSOD will compare the attributes based on their assessment results. For example, based on the analysis result, the attribute of Nama_Pemenang has a moderate cost and moderate benefit at the same time. This assessment shows that the Nama_Pemenang attribute has a clear utility for individuals, the community, and other organisations. However, this attribute has a severe adverse effect on individuals, the community, and other organisations.

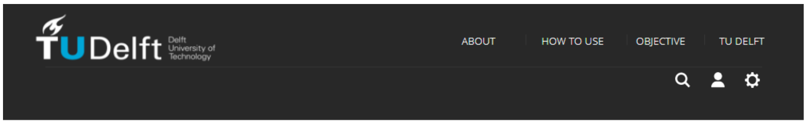


ATTRIBUTE	DISADVANTAGES	ADVANTAGES
NAMA PEMENANG	62 MODERATE	68 MODERATE
NPWP PEMENANG	55 MODERATE	57 MODERATE
NAMA AGENCY	57 MODERATE	55 MODERATE
NAMA PANITIA	60 MODERATE	58 MODERATE
NAMA SATKER	57 MODERATE	56 MODERATE
NAMA KABUPATEN	35 LOW	66 MODERATE
KODE PEMENANG	45 LOW	57 MODERATE
KODE LPSE	38 LOW	56 MODERATE
PAGU	37 LOW	50 MODERATE
NAMA PAKET	41 LOW	44 LOW
NILAI KONTRAK	44 LOW	56 MODERATE
HPS	44 LOW	47 MODERATE
SKOR PEMENANG	49 LOW	43 MODERATE
SKOR TOTAL	44 LOW	46 MODERATE

Figure 6-29 Construct matrix comparison

6.4.3.1.3 Construct pairwise comparison matrix

After the development of the comparison matrix in the previous step, the DSOD will construct a pairwise comparison matrix to provide the decision recommendation, as presented in Figure 6-30. For example, the attribute Nama_Pemenang has quantified as 62% to possible moderate disadvantage, and it has quantified as 68% to potential moderate benefit at the same time. Based on the qualitative values pairwise comparison matrix shown in Table 6-2, the DSOD recommends setting an additional screening to the dataset. The additional screening refers to releasing this dataset presents a high privacy disadvantage, and the benefits could outweigh the potential privacy disadvantages, or releasing this dataset presents a privacy disadvantage, and the potential benefits do not outweigh the potential privacy disadvantages.



Pairwise Comparison

ATTRIBUTE	DISADVANTAGE	ADVANTAGE	RECOMMENDATION
NAMA PEMENANG	62 MODERATE	68 MODERATE	ADDITIONAL SCREENING
NPWP PEMENANG	55 MODERATE	57 MODERATE	ADDITIONAL SCREENING
NAMA AGENCY	57 MODERATE	55 MODERATE	ADDITIONAL SCREENING
NAMA PANITIA	60 MODERATE	58 MODERATE	ADDITIONAL SCREENING
NAMA SATKER	57 MODERATE	56 MODERATE	ADDITIONAL SCREENING
NAMA KABUPATEN	35 LOW	66 MODERATE	OPEN
KODE PEMENANG	45 LOW	57 MODERATE	OPEN
KODE LPSE	38 LOW	56 MODERATE	OPEN
PAGU	37 LOW	50 MODERATE	OPEN
NAMA PAKET	41 LOW	44 LOW	ADDITIONAL SCREENING
NILAI KONTRAK	44 LOW	56 MODERATE	OPEN
HPS	44 LOW	47 MODERATE	OPEN
SKOR PEMENANG	49 LOW	43 MODERATE	OPEN
SKOR TOTAL	44 LOW	46 MODERATE	OPEN

Figure 6-30 Construct pairwise comparison matrix

6.4.3.2 The use of FMCDM in the prototype

To describe how the FMCDM approach works, we use a decision-making process consisting of three main phases: data source, evaluation, and decision. The entire process starts with selecting the dataset from the data source to create the input for the evaluation phase. The input data are processed next in the evaluation phase. The output of the evaluation, namely the decision stage, is a suggestion to make a decision. The latter is done by showing the rank of decision priority.

1) **Data Source**: First, we selected the type of dataset to define the criteria and sub-criteria of the disadvantages and benefits of opening data. This study applied four criteria and eight sub-criteria of the disadvantages and benefits as the input data.
2) **Evaluation**: In the second stage, we used FMCDM to assess the alternatives based on criteria defined in the data source elicitation phase, and the criteria use linguistic matrix values reflected by the Fuzzy. FMCDM works on the Fuzzy AHP technique has an essential role in measuring the relative importance of defined criteria for dealing with decision-making problems. To quantify the relative importance of the

disadvantages and advantages, we used expert judgment input. There are two main steps to conduct an evaluation process by the experts in AHP (Hancerliogullari et al., 2017; Podvezko, 2011): First, experts should rank the criteria in a descending or ascending order of their significance. Then, determining the most important criteria and compare the criteria and sub-criteria - with each other.

3) **Decision**: Finally, the outcome of this flow process is to get the best alternative's final weights as the priority of a decision.

6.4.3.2.1 Define Fuzzy linguistic scale

Construct a pairwise comparison matrix and evaluate the relative importance of criteria. The experts were asked to provide their consideration based on their knowledge and expertise. For simplicity, in this illustration, a pairwise comparison matrix for expert one is given in Figure 6-31**.** Before the experts started to quantify the criteria, we constructed a Fuzzy evaluation linguistic scale for the weights presented in Figure 6-31.

We created an intensity scale of importance between two elements to help the decision-maker or data analyst assess the pair-wise comparisons. The suggested numbers express a degree of preference between the two elements, as shown in Table 6-6.

Table 6-6 The fundamental scale for pair-wise comparison (Podvezko, 2011)

Intensity of importance	Definition	Explanation
1	Equal important (EI)	Two activities contribute equally to the objective
3	Weakly important (WI)	Experience and judgment slightly favour one activity over another
5	Essentially important (EI)	Experience and judgment strongly favour one activity over another
7	Very strongly important (VI)	An activity is favoured very strongly over another; its dominance demonstrated in practice
9	Absolutely important (AI)	The evidence favouring one activity over another is of the highest possible order of affirmation
2, 4, 6, 8	For a compromise between the above values	Sometimes one needs to interpolate a compromise judgment numerically because there is no good word to describe it
Reverse of above	If activity i has one of the above nonzero numbers assigned to it when compared with activity j, then j has the reciprocal value when compared with i	A comparison mandated by choosing the smaller element as the unit to estimate the larger one as a multiple of that unit

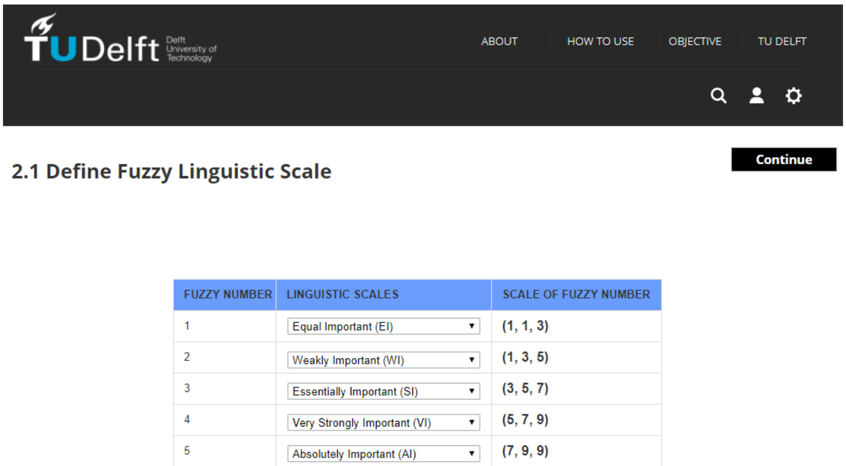
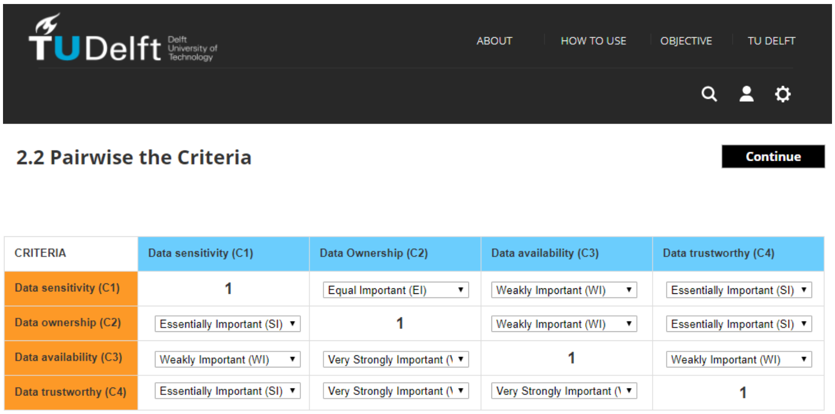


Figure 6-31 Define Fuzzy linguistic scale

6.4.3.2.2 Pairwise comparison of the criteria

After setting the scale for pairwise comparison in the fuzzy linguistic scale step, the decision-makers or data analyst then scale the pairwise criteria. In this step, we compare one criterion to another. For example, data sensitivity is equally as important as data ownership. Based on the expert's opinion, they qualitatively believe that the data sensitivity issue is essential than the selected dataset's data ownership. For instance, another example shows the experts qualitatively confident that data ownership is much more critical or fundamentally important than data sensitivity, as can be seen in Figure 6-32.



CRITERIA	Data sensitivity (C1)	Data Ownership (C2)	Data availability (C3)	Data trustworthy (C4)
Data sensitivity (C1)	1	Equal Important (EI)	Weakly Important (WI)	Essentially Important (SI)
Data ownership (C2)	Essentially Important (SI)	1	Weakly Important (WI)	Essentially Important (SI)
Data availability (C3)	Weakly Important (WI)	Very Strongly Important (	1	Weakly Important (WI)
Data trustworthy (C4)	Essentially Important (SI)	Very Strongly Important (	Very Strongly Important (	1

Figure 6-32 Pairwise the advantages and disadvantages criteria

6.4.3.2.3 Pairwise comparison of the sub-criteria

Furthermore, the next step is that the users then scale the pairwise sub-criteria. In this step, we compare one sub-criterion to another. For example, individual life-threatening in data sensitivity criteria are less important than data identifiable. It means that based on the expert's opinion, they qualitatively believe that an identifiable data issue is essential rather than the individual life-threatening from the case of the selected dataset. Moreover, the experts are confident that data recoverability is much more important than manageability data. These illustrations can be seen in Figure 6-33.

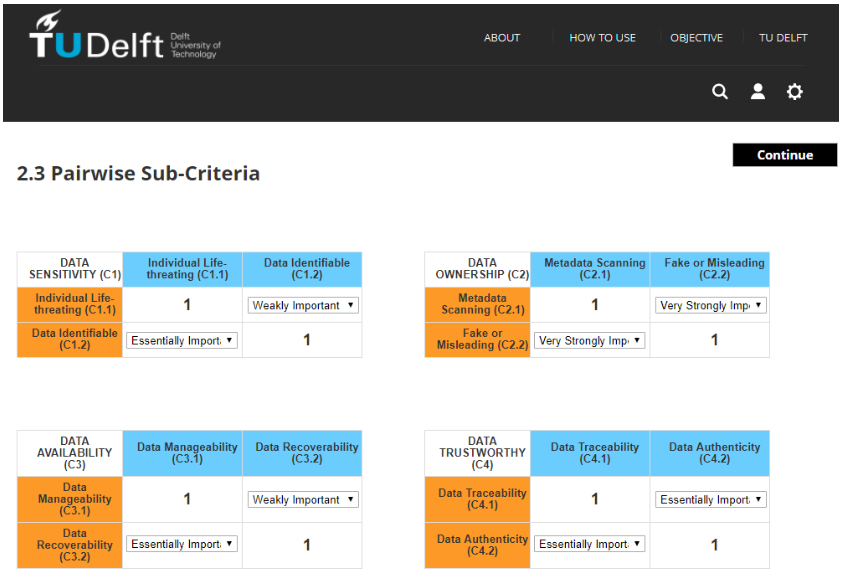


Figure 6-33 Pairwise advantages and disadvantages sub-criteria

6.4.3.2.4 Calculate Fuzzy weights matrix

The final weights of the alternatives are calculated using Eq. (3), (4), and (5). The linguistic terms provided in Figure 6-33 are utilised for the evaluation, and fuzzy operational laws were used for the calculation (Hancerliogullari et al., 2017; Hsieh et al., 2004). Illustrative examples for weights of sub-criteria C11 and C12 are given. Calculating sub-criteria: Linguistic terms for the pairwise comparison from Figure 6-34 and the corresponding fuzzy numbers. For example, the pairwise comparison of (C1.1 C1.2) is "Equal Important," and the fuzzy number of this linguistic term is (1,1,3).

$$\tilde{r}_{c11} = (\tilde{c}_{c11c11} \otimes \tilde{c}_{c11c12})^{\frac{1}{2}}$$

$$\tilde{r}_{c11} = \left((1,1,1) \otimes (3,5,7)\right)^{\frac{1}{2}}$$

$$\tilde{r}_{c11} = (1.73,2.23,2.64)$$

$$\tilde{r}_{c12} = (\tilde{c}_{c12c11} \otimes \tilde{c}_{c12c12})^{\frac{1}{2}}$$

$$\tilde{r}_{c12} = \left((1/(3,5,7)) \otimes (1,1,1)\right)^{\frac{1}{2}}$$

$$\tilde{r}_{c12} = (0.37,0.44,0.57)$$

Calculating weights: For calculating weights, we use eq. 4. In the previous step, we got the value of $\tilde{r}_{c1.1}$ and $\tilde{r}_{c1.2}$ and put these values in the following equation.

$$\tilde{w}_{c1.1} = (0.36,0.5,1.10)$$
$$\tilde{w}_{c1.2} = \tilde{r}_{c1.2} \otimes (\tilde{r}_{c1.1} + \tilde{r}_{c1.2})^{-1}$$
$$\tilde{w}_{c1.2} = (0.57,1,1) \otimes [(1,1,1.73) + (0.57,1,1)]^{-1}$$
$$\tilde{w}_{c1.2} = (0.2,0.5,0.63)$$

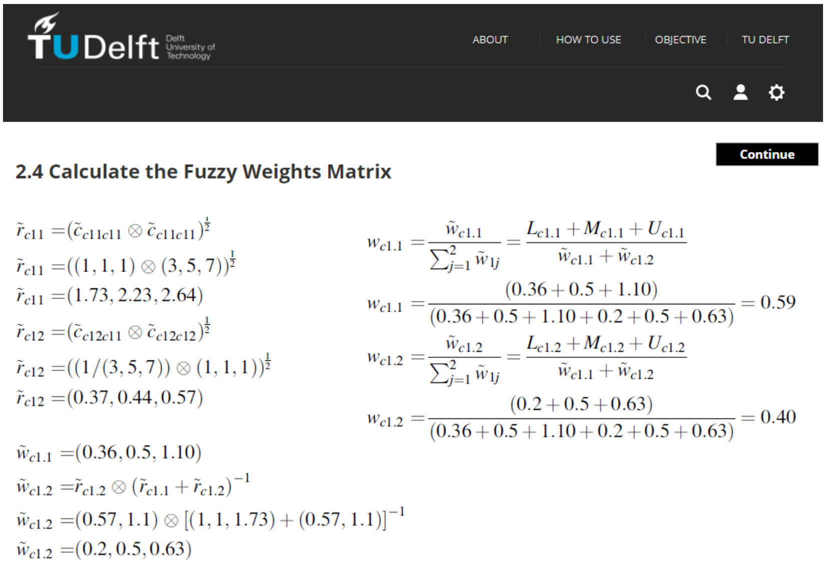


Figure 6-34 Calculate Fuzzy weights matrix

6.4.3.3 The use of Decision Tree Analysis

6.4.3.3.1 Assign payoffs table

The DTA employs numerical values to the probabilities in this step, including the action-taking place and the expected investment value. The assigned payoffs represent each combination's outcome in a table. This table uses costs terminology that describes the negative impact of a decision, like value for the expense and potential lost revenue (Adina Tofan, 2015; Delgado-Gómez et al., 2019).

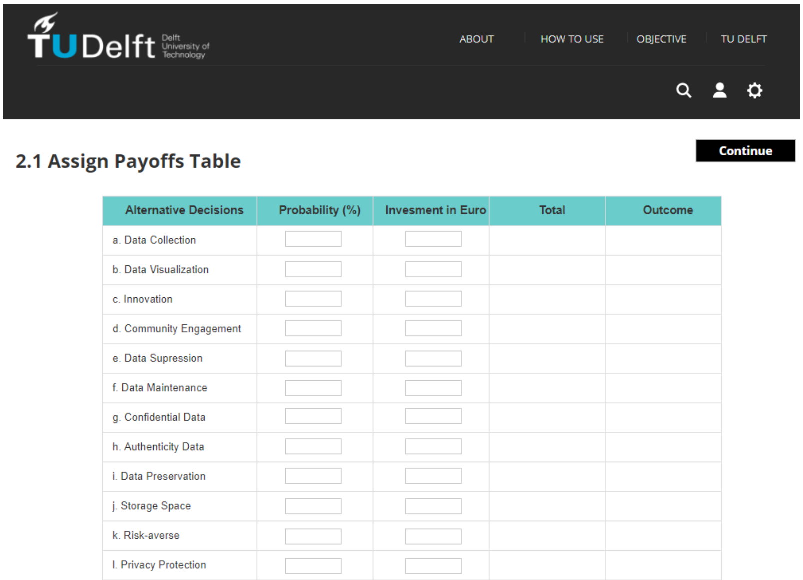


Figure 6-35 Assign payoffs table

6.4.3.3.2 Construct possible consequence

The process of constructing possible consequences is shown in Figure 6-36, resulting in the payoff results derived from the experts (as shown in Figure 6-35). The disadvantages and advantages of the three decision nodes (open, limited access, and closed) are compared from the constructed data. The numbered of each sub-element indicate the prediction of money expenses. The illustration of the construction of possible consequences in weighing the cost and benefits of opening data can be seen in Figure 6-36. The decision-makers can use this to select a decision.

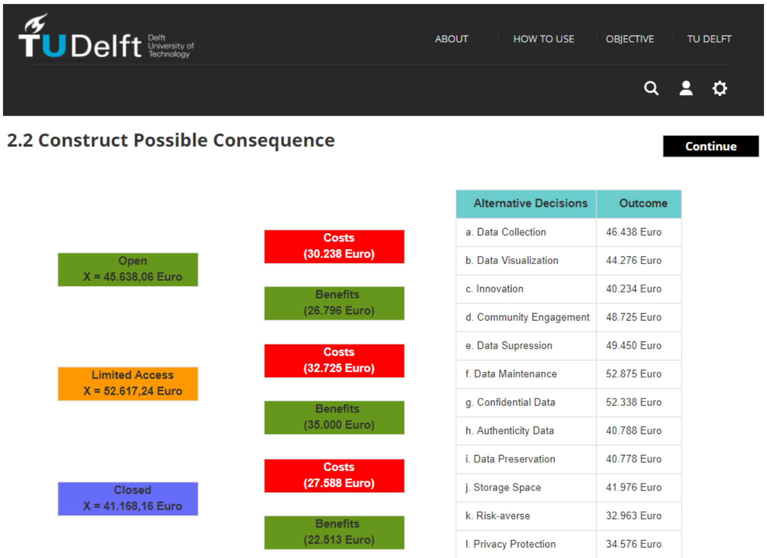


Figure 6-36 Construct a possible consequence

6.4.4 Decision-making

We already mentioned that there are four possible decisions to release the datasets. Table 6-7 shows a possible relationship between the advantages and disadvantages is shown, including some suggestions for which decision might be appropriate. Indeed, the actual decision is dependent on more factors, and this example should be viewed as merely an example. *Open decision* refers to the publishing of the dataset. It is recommended that the opening of the dataset has no or limited disadvantage to the individual or organizational identification (or other issues such as sensitivity), or the potential benefits of the dataset substantially outweigh the potential disadvantages. The *Limited access* decision refers to when the dataset is only released to a selected group of persons who adhere to some agreements to avoid the disadvantages. *Additional treatment decision* refers to reducing the disadvantages first before the dataset can be opened. Typically, this is decided when there might be significant disadvantages, but the opening of the datasets can bring much value. Finally, a *closed decision* is recommended if releasing the dataset generates a high or very high disadvantage to the individual or organization and significantly outweighs the potential benefits. Indeed, users can change how the recommendations are made in the DDOS.

Table 6-7 Example of weighing the advantage against the disadvantage

Potential occurs of benefits	Potential occur of disadvantages and their consequences				
	Very high disadvantage	**High disadvantage**	**Moderate disadvantage**	**Low disadvantage**	**Very low disadvantage**
Very high advantage	Additional treatment	Additional treatment	Limited access	Open	Open
High advantage	Additional treatment	Additional treatment	Limited access	Limited access	Open
Moderate advantage	Closed	Additional treatment	Additional treatment	Limited access	Limited access
Low advantage	Closed	Closed	Additional treatment	Additional treatment	Limited access
Very low advantage	Closed	Closed	Closed	Additional treatment	Additional treatment

6.4.4.1 The use of Bayesian-belief Networks

In this step, the DSOD provides alternative decision possibilities and recommendations to the decision-makers (see Figure 6-37). The four possible decisions shown in the histogram contain the scores of the decision status. Figure 6-37 presents the quantification resulting from the decision structure (see Figure 6-26). The bar with different colours represents the decision recommendation based on the weighing process between the advantages and disadvantages of the dataset attribute. For example, the yellow-coloured attribute name of nama_pemenang represents the additional screening decision, whereas the nama_kabupaten attribute having the blue colour indicates the open decision. The result shows that the DSOD generated six attributes with additional screening decisions, e.g., nama_pemenang, NPWP_pemenang, nama_agency, nama_satker, and nama_paket. At the same time, the DSOD also supports a recommendation for another eight attributes for opening decision, e.g., nama_kabupaten, kode_pemenang, kode_LPSE, pagu, nilai_kontrak, HPS, skor_pemenang, and skor_total. Based on this result, the decision-makers can select alternative treatment methods as recommendations (see Table 6-8) with the privacy and utility impact.

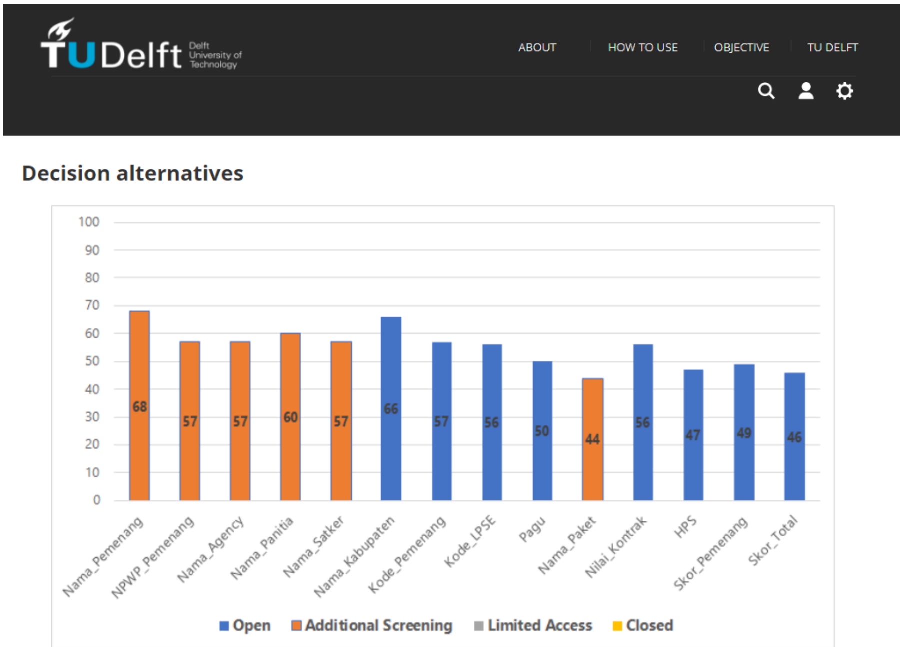


Figure 6-37 Decision alternatives in Bayesian-belief Networks

Next, the DOSD provides support for deciding about alternative treatments to reduce the disadvantages of the datasets. There are five possible alternative treatments, namely suppression, blurring, pseudonymization, aggregation, and visualization of the data, as shown in Table 6-8. Each treatment contains a description, the impact on privacy and utility. Indeed, new treatment plans can be added.

Table 6-8 The treatment plans proposed in using BbN method

	Treatment	Description	Privacy Impact	Utility impact
1	Suppression	Remove a data field or an individual record to prevent individuals' identification in small groups or those with unique characteristics.	Removing the field removes the risk created by those fields and lowers the likelihood of linking one dataset to another based on that information. Removing individual records can also effectively protect the privacy of those individuals. Suppression cannot guarantee absolute privacy because there is always a chance that the remaining data can be re-identified using an auxiliary dataset.	This approach removes all utility added by the suppressed field or record, skews the results, or gives false impressions about the underlying data.
2	Blurring	Reduce the precision of disclosed data to minimise individual identification certainties by replacing precise data values with ranges or sets.	The more specific a data value is, the easier it will generally be to single out an individual. However, even relatively broad categories cannot guarantee absolute privacy because there is always a chance that the remaining data can be re-identified using an auxiliary dataset.	Generalising data fields can render data useless for more granular analysis, skew results slightly, or give false impressions about the underlying data.

	Treatment	Description	Privacy Impact	Utility impact
3	Pseudonymization	Replace direct identifiers with a pseudonym (such as a randomly generated value, an encrypted identifier, or a statistical linkage key).	Pseudonymization refers to removing the association between an individual and their data and substitute it with another key, lowering but not eliminating the risk of re-identification.	Pseudonymization can allow for information about an individual to be linked across multiple records, increasing its utility for a wide variety of purposes.
4	Aggregation	Summarise the data across the population and then release a report based on those data (such as contingency tables or summary statistics), rather than releasing individual-level data.	Aggregating data can effectively protect privacy as there is no raw data directly tied to an individual. However, experts recommend minimum cell sizes of 5-10 records.	Aggregation is more helpful in examining the performance of a group or cohort. Because the raw data is not presented, it cannot be relied on to generate additional insights.
5	Visualisation	Data may be presented in more privacy-protective formats rather than providing users access to raw microdata, such as data visualisations or heat maps.	When data is released in non-tabular formats, individual data records are typically more obscure and harder to link to other auxiliary datasets, protecting individual privacy.	Data released in these sorts of formats may still be highly useful for various purposes, although not all. These formats may also limit how datasets can be combined or built on to generate new insights.

6.4.4.2 The use of Fuzzy Multi-criteria Decision Making

Also, with FmcDM, the next step is to make decisions about the opening of datasets. In contrast to BbN. A calculation approach is applied for all pairwise comparisons. The final weights of the alternatives are provided in Table 6-3. An illustrative example of W_{A1} is given as follows:

$$W_{A1} = C1 \times C11 \times A1 + C1 \times C12 \times A1 + \cdots + C4 \times C41 \times A1 + C4 \times C42 \times A1$$

$$W_{A1} = 0.53 \times 0.59 \times 0.39 + 0.53 \times 0.40 \times 0.41 + \cdots + 0.07 \times 0.59 \times 0.44 + 0.07 \times 0.40 \times 0.35$$

$$W_{A1} = 0.34$$

Table 6-9 Final weights of the criteria and alternatives

Decision alternative	C1		C2		C3		C4		
	0.53		0.25		0.13		0.07		
	C1.1	C1.2	C2.1	C2.2	C.3.1	C3.2	C4.1	C4.2	
	0.59	0.40	0.82	0.17	0.59	0.40	0.59	0.40	Weight
A1 (Open)	0.39	0.41	0.41	0.41	0.44	0.35	0.44	0.35	0.34
A2 (Additional Screening)	0.40	0.39	0.82	0.83	0.23	0.44	0.23	0.44	0.43
A3 (Limited Access)	0.06	0.13	0.26	0.13	0.08	0.15	0.08	0.15	0.08
A4 (Closed)	0.05	0.05	0.10	0.05	0.22	0.05	0.22	0.05	0.06

According to Table 6-9, the highest-ranked of the decision alternative to open data is A2 (additional screening) with 0.43 score, followed by A1 (open) with 0.34 score and A3 (limited access) with 0.08 score. In comparison, the least ranked decision recommendation is A4 (closed), with 0.06 score.

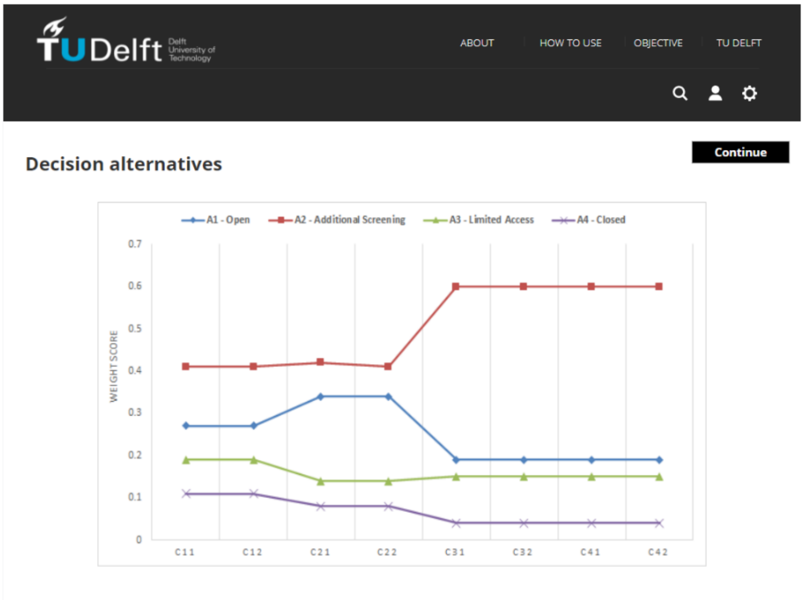


Figure 6-38 Decision alternatives

After defining the priority of the decision, each decision can be ranked. The mechanism of how to rank the decision is based on the expert's input. For example, the attribute of Nama_Satker is ranked in order as (1) limited access, (2) additional screening, (3) closed, and (4) open. Based on this rank, the highest-ranked decision for the Nama_Satker attribute is giving limited access to this field.

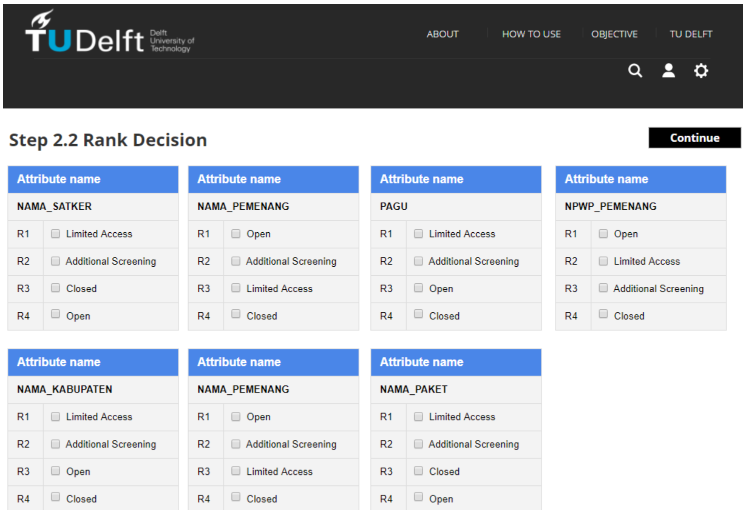


Figure 6-39 Rank Decision alternatives

To present the recommendations based on the analysis process's final results using FMCDM, we designed a graphical view to support decision-makers in deciding to release their dataset. Figure 6-39 shows how the Fuzzy AHP could help decision-makers better understand the decision-making's action plans. In a similar vein as with BbN, decisions can be made.

6.4.4.3 The use of Decision Tree Analysis

The final step in developing DTA is to decide and provide some recommendations presented in decision action plans. To provide the most suitable decision between the three alternatives (open, limited access, and closed) to the decision-makers, weighing the costs and benefits is used. Next, from the Expected Monetary Value (EMV) results, the DTA can recommend a decision to the highest priority that might influence institutional revenue streams' investment. We classify the findings of the study into two parts, namely:

Possible paths and pay-offs are shown in a DTA, including their chances, as can be seen in Figure 6-40. Every decision alternatives provide the estimation of payoffs in the Euro currency. For this illustration, the highest investment for the costs factor in the open data domain is data maintenance. Data maintenance, in this case, is the sub-nodes of the limited access decision. In this example, the highest potential benefit requires 52,000 Euros investment and would be a net benefit for the government institutions. In this case, the limited access decision can potentially have high costs and result in high new revenues.

ABOUT | HOW TO USE | OBJECTIVE | TU DELFT

3.1 Possible Paths and Total Payoffs

Continue

Terminal	Total Payoff
Decision → Open → Decision 1 → Costs → Chance 1 → Data collection	46.438 Euro
Decision → Open → Decision 1 → Costs → Chance 1 → Data visualization	44.276 Euro
Decision → Open → Decision 1 → Benefits → Chance 2 → New knowledge	40.234 Euro
Decision → Open → Decision 1 → Benefits → Chance 2 → Community engagement	40.335 Euro
Decision → Limited access → Decision 2 → Costs → Chance 3 → Data suppression	48.725 Euro
Decision → Limited access → Decision 2 → Costs → Chance 3 → Data maintenance	49.450 Euro
Decision → Limited access → Decision 2 → Benefits → Chance 4 → Confidential data	52.875 Euro
Decision → Limited access → Decision 2 → Benefits → Chance 4 → Authentication data	52.338 Euro
Decision → Closed → Decision 3 → Costs → Chance 5 → Data preservation	40.788 Euro
Decision → Closed → Decision 3 → Costs → Chance 5 → Storage space	41.976 Euro
Decision → Closed → Decision 3 → Benefits → Chance 6 → Risks-averse	32.963 Euro
Decision → Closed → Decision 3 → Benefits → Chance 6 → Privacy protection	34.576 Euro

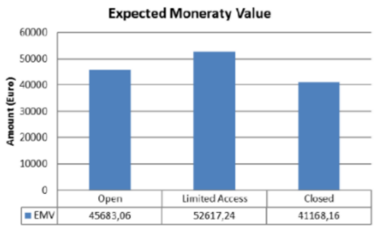
Figure 6-40 Possible paths and total payoffs

The EMV resulted from the DTA shows that the limited access decision gains the highest monetary value. The EMV of each decision is derived from the probability-weighted average of the expected outcome. Figure 6-41 presents the details illustrating possible EMV results and ranges of the potential investment.

ABOUT | HOW TO USE | OBJECTIVE | TU DELFT

3.2 Final Decision

Continue



Decision	EMV	Range
Open	45.638,06 Euro	44.276 – 46.438 Euro
Limited Access	52.617,24 Euro	52.338 – 52.875 Euro
Closed	41.168,16 Euro	40.788 – 41.976 Euro

Figure 6-41 Final decision in using DTA

In the next step, there is a process named "Reiterate". In the case that the decision-makers require to re-analyse the dataset, the DSOD can iterate the process from the beginning or re-process the dataset by starting at a certain step. The decision-makers can use the DSOD to modify, refine, and evaluate the selected dataset before sending it back to the cycle process. The dataset can be modified in the iterations, and the number of the selected attributes can be re-analysed.

6.5 Prototype validation and testing

Prototype validation aims to show that a developed system both conforms to its specification and meets the users' requirements (Sommerville, 2011). Prototype validation includes prototyping testing and reviewing the process from the user requirements definition to program development (Carr & Verner, 1998; Sommerville, 2011). The process was already reviewed in the previous section, where this section focused on prototype testing.

Prototyping *testing* is a phase in the testing process in which stakeholders or system users provide input and feedback on the system testing (Bernstein, 1996; Pliskin & Shoval, 1987; Sommerville, 2011). In this research, we use three different types of the user's prototype testing introduced by (Sommerville, 2011), namely alpha, beta and acceptance testing.

6.5.1 Alpha testing

In alpha testing, the decision-makers and data analysts work with the researcher to test the system at the developer's site. This testing type aims to identify problems and issues that are not readily tangible during the DSOD system development. For our research, we thoroughly considered the requirements of the proposed DSOD in collaboration with 7 PhD students.

6.5.2 Beta testing

In beta testing, a release version of the developed prototype is made available to users or stakeholders to experiment and raise problems that they discover with the system developers (Sommerville, 2011). Usually, beta testing occurs when an early or unfinished version of the prototype is made available to end-users to give attention in an evaluation (Smith, 1991; Sommerville, 2011). The users of the beta tester participants were derived from a selected group of potential users who are early adopters of the developed prototype. Because we use an evolutionary prototyping approach, we organised three different user groups consisted of 49 participants in total to test the different releases of the DSOD prototype in this research. The first beta tests involved a group from academia consisting of 15 participants, the associate professors, PhD candidates, bachelor's and master's students, and independent researchers from Indonesia. We derived 18 participants from local government institutions in the second beta tests, including decision-makers, policy-makers, civil servants, politicians, and data analysts. The third beta test involved a group of 16 participants from the community, including professional open government data analysts and other non-governmental organisations.

6.5.1 Acceptance testing

Acceptance testing aims to test a system to decide whether the developed prototype is ready to be accepted by the users to be deployed in the selected environment (Sommerville, 2011). Acceptance testing takes place after releasing alpha and beta testing. It involves a user formally testing a system to decide whether it should be accepted by the system developer (Smith, 1991; Sommerville, 2011). In this research, we used an acceptance testing protocol (see Appendix B, Part III – Acceptance of the DSOD).

6.6 Conclusions

This chapter addressed the third design science research process, namely design and developing the decision-making support to open data. In this research, we presented our prototype approach by employing five main steps: define the objective, use the evolutionary prototyping, develop the functionalities of the DSOD, construct the DSOD prototype, and validate the DSOD prototype. The prototype provides support for the three methods. As the methods are different, they needed to be implemented differently, although we tried to have the user experience similarly. The latter should enable comparison between the methods without being influenced by differences in the user interface.

Chapter 7
Quasi-Experiment using the DSOD

This chapter discusses the fourth phase of this research, namely the evaluation of the three decision-making methods. This chapter's main objective is to answer the fourth research question (RQ#4): *What are the differences between BbN, FMCDM, and DTA to support decision-making about opening the dataset?* To address the aim of the research question, we decomposed this chapter into five sub-sections. Section 7-1 presents the evaluation methodology used in this research, including the quasi-experimental approach, quasi-experiment structure, and quasi-experiment preparation. Second, we present quasi-experiment 1 (e-procurement), followed by quasi-experiment 2 (medical records) in Section 7.3. Next, we provide a comparative analysis of two quasi-experiments. Finally, we draw conclusions and answer the fourth research question.

In this study, we use two quasi-experiments to evaluate the three different methods. The first empirical setting is based on the e-procurement case study. This case was performed with participants of the Indonesian government. An electronic procurement system (e-procurement) is the electronic-based processing of the transaction related to the purchase orders (Boer, Harink, & Heijboer, 2002; Klabi et al., 2018). The use of e-procurement systems can promote the effectiveness and efficiency of purchasing procedures, simplify administration, and improve public transparency (Boer et al., 2002; Czarnitzkia, Hünermund, & Moshgbar, 2020). The government wants to open data about e-procurement projects. However, publishing the e-procurement dataset encounters several risks. The potential disadvantages encountered from the e-procurement dataset's opening are the possibility of personal and company privacy violations, opening inaccurate procurement data, and contradicting or against the law.

This research's second empirical setting is based on the medical records case study, which is part of the Indonesian government. The terms of health records, or medical records, are used to describe the systematic documentation of a single patient's medical history and care across time within one specific health care provider's authority (Spooner & Pesaturo, 2013). The medical records include various archives entered over time by healthcare professionals, recording observations and administration of therapies, test results, x-ray photographs, and psychology assessment records. On the one hand, the analysis of medical records may create societal benefits by monitoring the current situation and identifying trends. However, disclosing the medical record datasets can result in several potential disadvantages, such as revealing the patients' identity, misusing the patients' medical history, and disclosing patient information without proper authorization.

7.1 Evaluation methodology

Evaluation is an essential component of the research process to demonstrate the design artefact's utility, quality, and efficacy (Hevner et al., 2004; Matthews & Wensveen, 2015). In this section, the evaluation method consists of two main parts. First, we present the quasi-experimental approach used to conduct the experiments in a highly controlled environment and compare the case study's participants. Second, we present the quasi-experiment structure, including pre-test, post-test, and performing scenarios. Third, we provide the quasi-experiment preparation, including the statistical software and the reliability analysis used in the pre-test and post-test quasi-experiment.

7.1.1 Quasi-experimental approach

Experimental studies are conducted to determine the cause and effect of a treatment, program, or other implementation (Champbell & Stanley, 1963; Thyer, 2012). The quasi-experimental approach is often utilised to evaluate various ways to improve the present situation at hand (Champbell & Stanley, 1963). Most experiments are conducted in a highly controlled environment, such as the laboratory, whereby a random sample of test participants has been selected prior. They are usually conducted as a comparison test between at least two groups of participants, a treatment group and a control group (Champbell & Stanley, 1963). The control group will receive the standardised condition, while the treatment group is those who will receive the treatment.

In this research, the DSR approach aims to evaluate the three different methods, BbN, FMCDM, and DTA. Section 5.4, a comparison of methods, shows that each method has different benefits and objectives (Ceballos et al., 2017; Heckerman, 2008; S. Zhang, 2012). The advantage of using the BbN method is that the result is more accurate and can encounter the uncertainties and possible consequences. The benefit of using FMCDM is the ability to systematically construct a hierarchy of decisions, including its alternative and ranking them into the best options. Nevertheless, The FMCDM method is relatively difficult to comprehend, and the model interpretation is not easy and, like BbN, it also needs an understanding of mathematics. Furthermore, DTA is the least time-consuming to analyse the dataset. However, the DTA is challenging, and the re-development of the decision hierarchies is time-consuming when decision-makers modify variables during the analysis process.

A quasi-experiment was conducted by applying a two-group random assignment pre-test and post-test design. There are four main evaluation factors to measure the effects of the developed prototype in this study. The evaluation factors are how transparent is the process, how accurate is the expected results, how easy it is to understand the steps of the decision-making process, and how efficient the time is to process the proposed DSOD.

7.1.2 Structure of the quasi-experiments

In this research, we designed the quasi-experiments structure using five main steps. First, we introduced the motivation and objective of the experimental case study. Thereafter, we provided the introduction of the proposed DSOD to the participants. Second, we conducted a pre-test by asking the demographics and their experiences in using open government data. Third, we performed scenarios in random sequence methods. In this step, the three methods, namely BbN, FMCDM, and DTA, were used. Fourth, we conducted a post-test by asking the participants to fill the questionnaire to evaluate the methods. In this way, the methods were evaluated based on their transparency, accuracy, perceived ease of use, usefulness, and acceptance of the DOSD. Figure 7-1 illustrates the structure and timeline to perform the experiment.

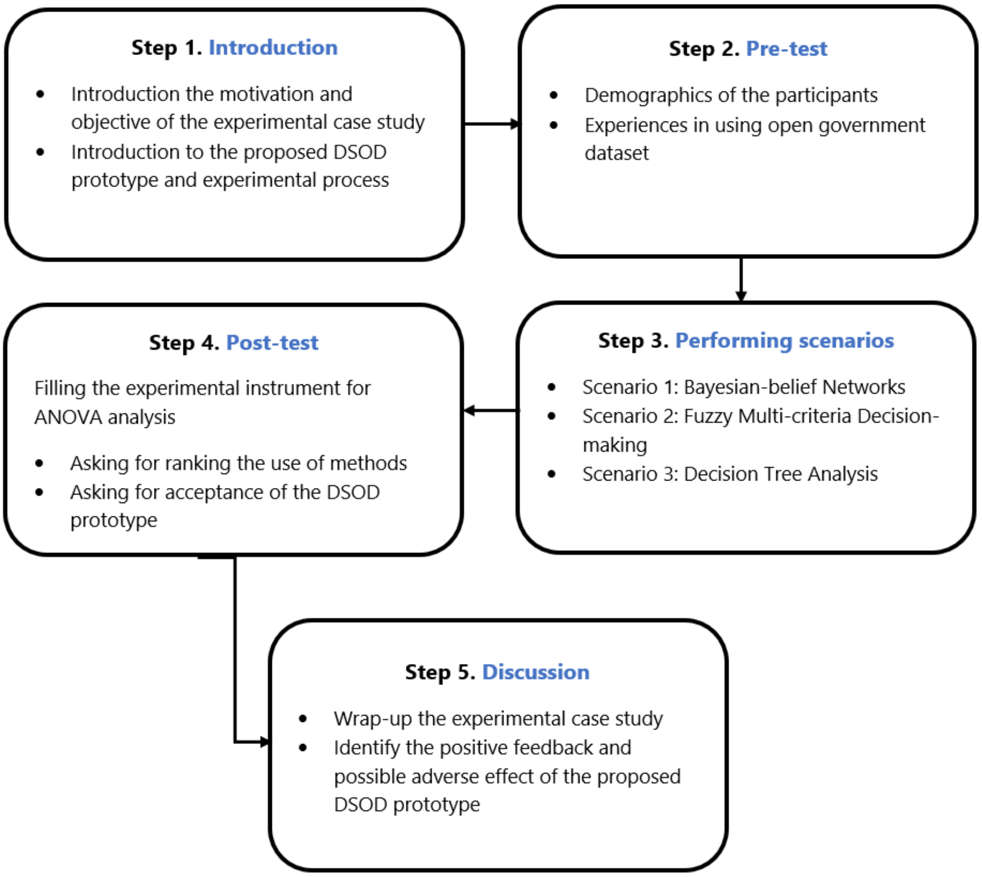


Figure 7-1 Structure and timeline of the experiment

7.1.3 Preparation of the Quasi-experiment

In this research, we use the IBM SPSS Statistics software version 23 to analyse the collected data. Reliability analysis was conducted to measure the constructs' consistency in the pre-test and post-test quasi-experiment instrument as part of the data preparation. Reliability refers to the degree to which measures are free from error and yield consistent results (Peter, 1979). We use Cronbach's Alpha to calculate and obtain information about the constructs' reliability to define the reliability coefficient. Murphy and Davidshofer (1988) state that alpha values below 0.6 are unacceptable, values of 0.7 are low, values between 0.8 and 0.9 are moderate to high, and values around 0.9 are high. Others (e.g., Davis, 1964; Nunnally, 1967) have recommended a lower acceptance boundary and believe that Alpha values between 0.5 and 0.6 can still be acceptable. Based on the statistical reliability test using the IBM SPSS 23, the pre-test values in this study are between 0.793 to 0.832, which are acceptable values for Cronbach's alpha (Field, 2009). At the same time, we also analyse the post-test reliability, which was between 0.852 and 0.870. Table 7-2 shows Cronbach's alpha values for the five items used in our DSOD prototype.

Table 7-1 Reliability analysis of the construct of the pre-test and post-test

Quasi-experiment	The functionalities of the DSOD prototype	Number of items	Cronbach's Alpha
Pre-test	(1) Initialisation	7	0.793
	(2) Analysing the dataset	6	0.806
	(3) Weighing the dataset	5	0.813
	(4) Decision-making	4	0.832
	(5) Updating the dataset	4	0.819
Post-test	(1) Initialisation	7	0.852
	(2) Analysing the dataset	6	0.870
	(3) Weighing the dataset	5	0.860
	(4) Decision-making	4	0.865
	(5) Updating the dataset	4	0.856

7.1.4 MANOVA analysis to evaluate the three methods

In this study, we used a Two-way Multivariate Analysis of Variance (MANOVA) to evaluate three dependent variables (BbN, FMCDM, and DTA) in terms of three independent groups of participants (academia, government, and community). We used four most relevant from eight variables in total (see Appendix B Part III) to measure the effects on the developed prototype, e.g., (1) transparency of the process, (2) accuracy of the results, (3) perceived ease of use to understand the decision-making process's steps and (4) usefulness of the proposed DSOD. The use of three different methods (BbN, FMCDM, and DTA) was evaluated on these variables by asking the participants to fill in the survey.

MANOVA is used as an analytical tool to test whether there is a difference in the mean between groups (Nath & Pavur, 1985). MANOVA analysis is often utilised in experimental-based research, where there are several treatments. In this case, we wanted to test whether there were significant differences between these treatments. MANOVA is one of the various types of parametric tests because it requires a normal distribution of the dependent variable per treatment or a normal distribution of the residuals (Todorov & Filzmoser, 2010). This normality requirement assumes that the sample is taken randomly and can represent the entire population so that the research results can be used as generalisations (Todorov & Filzmoser, 2010; J. Zhang, 2012).

The two-way MANOVA aims to understand the interaction and significant difference between independent variables on two or more dependent variables (Tabachnick & Fidell, 2011). This study used three independent variables from three different groups of participants to evaluate the proposed DSOD, namely government, academia, and community. Whereas in the dependent variables, we used three different methods: BbN, FMCDM, and DTA.

MANOVA is used to perform multivariable comparative analysis. The comparative analysis technique using the "t" test, which is to look for a significant difference between the two means, is only effective when the number of variables is two (Nath & Pavur, 1985). MANOVA is used to compare population means, not population variants. The right type of data for MANOVA is nominal and ordinal in the independent variable. If the independent variable's data is in the form of intervals or ratios, it must be converted into an ordinal or nominal form. At the same time, the dependent variable is interval or ratio data.

The normality test is a test used to determine whether the data population is normally distributed or not and measure the data with ordinal, interval, or ratio scales (Kutner & Wasserman, 1996). Thus, for the data normality test analysis, the data must come from a normal distribution. If the variable is not normally distributed, the method used is non-parametric statistics. In addition, the type of data distribution can be determined from the characteristics of the data itself, and it can also be done by testing whether the data tends to a normal distribution. Normality is a distribution that shows a balanced distribution of data; most of the data are in the middle value. Normality is a must and the first requirement in parametric analysis and regression analysis. The normality test aims to test whether, in the regression model, the confounding or residual variables have a normal distribution. If this assumption is violated, then the statistical test will be invalid or biased, especially for small samples. The normality test can be done through two approaches, namely descriptive and inferential. For our situation, we tested if the variables were normally distributed.

Next, we conducted a descriptive analysis. Descriptive statistical analysis is a method for collecting, processing, simplifying, presenting, and analysing quantitative data descriptively to provide an orderly picture of an event (Stone, Bleibaum, & Thomas, 2020). This enabled us to compare the scores of the three independent variables (academia, government, and community) on the three dependent variables (BbN, FMCDM, and DTA).

7.2 Quasi-experiment 1: Deciding on opening e-procurement dataset

The first empirical setting of this research is based on the electronic procurement case study. E-procurement is the electronic-based transaction processing related to purchasing orders (Czarnitzkia et al., 2020). The use of e-procurement systems can promote the effectiveness and efficiency of purchasing procedures, simplify administration, and improve public transparency. Opening an electronic procurement dataset to the public can reap several benefits (Czarnitzkia et al., 2020). First, competition between private business institutions increases a government's opportunities for getting better value and uses public resources more efficiently. Second, making the application process to the available services more transparent and helps for fighting corrupt practices. Third, opening procurement contracts and reports to the public increase's legal certainty.

The opening of a procurement system should make it easier for the public to monitor development projects. For example, Indonesia Corruption Watch (ICW) organisation, in collaborating with Indonesia National Public Procurement Agency

(LKPP), wants to monitor the government's programs starting from project planning, including project value, which and how many participants, institutions, or organisations are involved in the bidding process, who is the owner of the project, how the bids are evaluated, and how performed the project is realized.

Nevertheless, the number of e-procurement datasets is enormous. However, the opening of these datasets might result in several disadvantages. The potential disadvantages are the possibility of personal privacy violation, the opening of companies' sensitive information, the opening of inaccurate procurement data, and might contradict or even be against the law. The public and community's capacity to monitor the electronic procurement systems is very limited, and its opening can yield many benefits. Therefore, we use this e-procurement dataset to help relevant stakeholders decide whether a dataset should be opened and pre-process the datasets before releasing them to the public.

Table 7-2 Characteristics of the group participants to conduct beta testing

	Quasi-experiment 1A	Quasi-experiment 1B	Quasi-experiment 1C
Date	29 July 2019	31 July 2019	8 August 2019
Duration	140 minutes	125 minutes	130 minutes
Number of participants	18	15	16
Type of group participants	Local government institution	Academia	Community
Location	Indonesia	Indonesia	Indonesia
Step 1. Introduction	✔	✔	✔
Step 2. Pre-test	✔	✔	✔
Step 3. Performing scenarios	✔	✔	✔
Step 4. Post-test	✔	✔	✔
Step 5. Discussion	✔	✔	✔

Table 7-2 shows the characteristics of the group to conduct a quasi-experiment in the e-procurement case study. The participants were asked to conduct the decision-making processes in the electronic procurement (e-procurement) case study. We used pre-test and post-test approaches to perform the quasi-experiments. The participants from the three different groups were selected non-randomly. The quasi-experiment was conducted in July and August 2019, Indonesia. The quasi-experiments took between 125 and 140 minutes in total to test the proposed DSOD prototype. With three different groups, the quasi-experiments were conducted, e.g., with government officials (experiment 1a), academia (experiment 1b) and community members (experiment 1b). We involved 18 local government officials from several related departments, including the division of public information disclosure. There were 15 participants from academia, e.g., associate professor and PhD candidates in the open data research area, lecturers, senior researchers, and bachelor and master's students from Telkom University Indonesia. The 16 participants in the community group included non-government organisations, such as Indonesia Corruption Watch (ICW) and Open Data Labs Jakarta. The participants from the communities included data analytics for analysing the potential fraud of the e-procurement system.

7.2.1 Demographics of the participants

This section discusses the characteristics and differences between the quasi-experiment participants. This study used six characteristics to represent the participants' demographics, including gender, age, educational level, organisation type, and current job function of the participants. In total, there were 49 participants actively involved in this quasi-experiment case study. In all three quasi-experiments, the male (65%) participants dominated. The percentage of age (see Figure 7-2) majority were participants ranging from 25 to 34 years old (35%). In contrast, the youngest participants were between 18 to 24 years old (9%), and the oldest participants were above 54 years old (5%). This demographic indicates that there were no significant differences between the three different groups of participants in age.

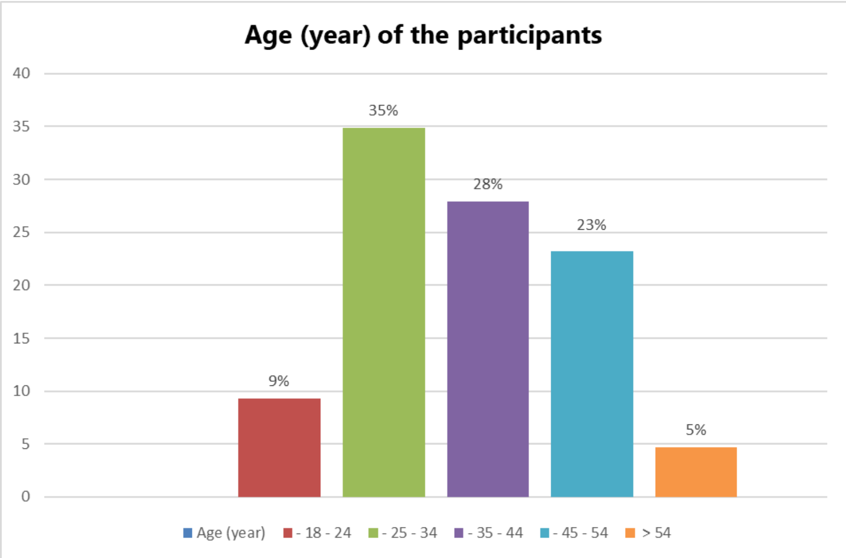


Figure 7-2 Distribution of the age of the participants

From the education background, the survey shows that most of the participants had a master degree level (58%), followed by the bachelor and doctorate level (23% and 19%, respectively). This result suggests that the participants have a good level of educational background in terms of their knowledge, expertise, and human cognition.

Figure 7-3 shows the type of organisations of the participants of the quasi-experiment. From the organisational types, our survey shows that most participants were working at local government institutions (31%), followed by the participants from university or academia and community-based organisations (23% and 22%, respectively). At the same time, other participant's organisations came from non-profit organisations, business and private organisations, and ministry or governmental departments (10%, 8%, and 6%, respectively).

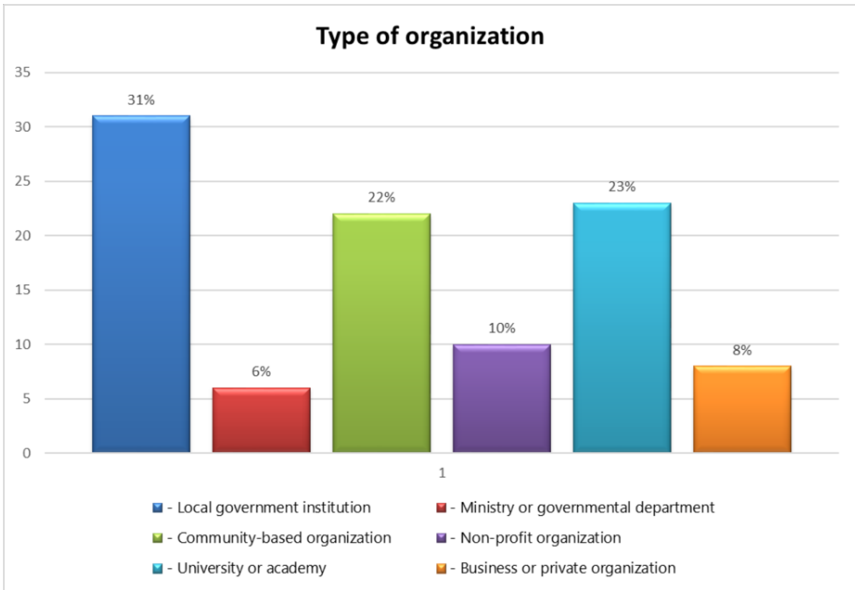


Figure 7-3 Type of organisation of the quasi-experiment participants

Figure 7-4 shows the job function of the participants. Most participants are responsible for analysing and investigating datasets from the OGD portal (28%). The participants' second job function was supervising those who provide services to the public, followed by coordinating or administering one or more programs (17% and 13%, respectively). Meanwhile, other job roles can be defined as administrative support (9%), technical services (8%), policy-making (7%), technical specialist (4%), and researchers (2%). From this result, our study found that the participant's job functions were varied. These results suggest that there were no significant differences between the three different groups of participants in job functions.

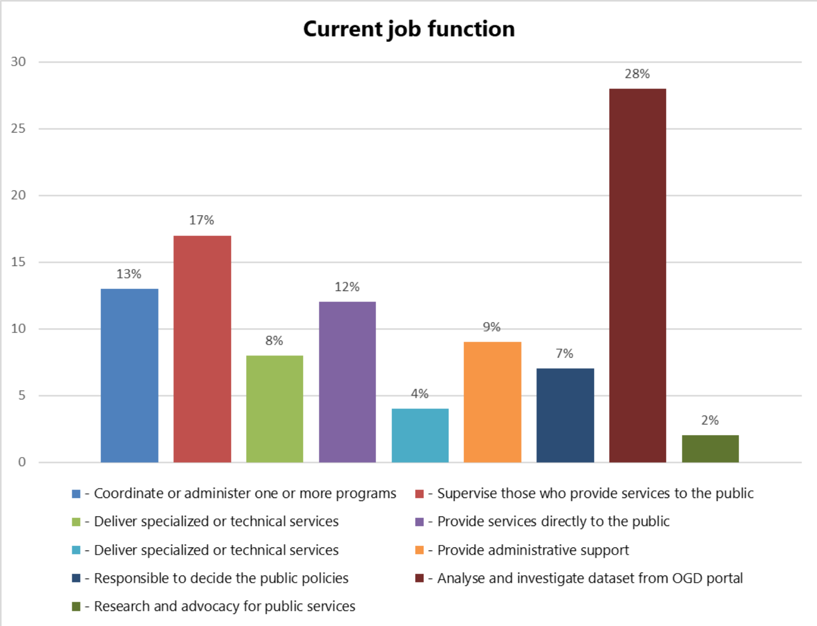


Figure 7-4 Current job function of the quasi-experiment participants

7.2.2 Experience in using open government data

This section elaborates on the participants' experience using open government data by looking at their years of experience and the actual use of opening data. This study used four characteristics to represent the participants' experience and expertise in using OGD datasets. First, the number of years of experience using ODD. Most participants' experience ranged between 6 to 10 years (54%), followed by between 1 to 5 years (30%), and some had between 11 to 15 years of experience (14%).

Figure 7-6 shows the frequency of the participants in using open government datasets. Half of the participants use OGD datasets daily or multiple times per day (56%), followed by the monthly use or a few times per month use of OGD datasets (21%). About 18% of the participants actively used OGD datasets each week or a few times per week, while about 5% of the participant rarely use the OGD dataset yearly or a few times per year. Overall this shows that the participants were well-experienced with OGD. There were no significant differences between the three different groups of participants in frequency in using the OGD dataset.

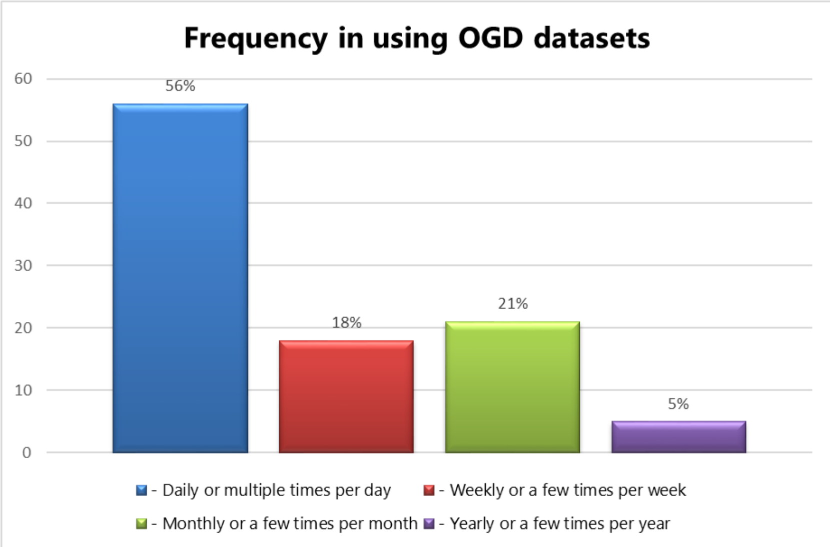


Figure 7-5 Frequency in using OGD dataset of the quasi-experiment

7.2.3 Measurement of the quasi-experiment

In this research, we used four variables to measure the effects on the developed prototype. The variables include (1) transparency of the process, (2) accuracy of the results, (3) easiness to understand the decision-making process's steps, and (4) efficiency of the use of the proposed DSOD.

The use of three different methods (BbN, FMCDM, and DTA) was evaluated because each method appears to have various advantages, as discussed in Chapter 5, Section 5.4.2 (Comparison of the Methods)-. A quasi-experiment setting was conducted by applying two non-random groups to perform a pre-test and post-test approach. The quasi-experimental design compares the four factors: transparency, accuracy, easiness, and efficiency before testing the DSOD prototype.

This study uses MANOVA to test the different participant groups for the three different methods (BbN, FMCDM, and DTA). The variance analysis aims to detect significant factors in a multi-factor model (Kutner & Wasserman, 1996). In this study, we designed the group of participants as the independent variable, and the methods represent the dependent variable.

7.2.3.1 Transparency of the DSOD

Process transparency refers to the transparency of every step involved in the decision-making process to analyse the advantages and disadvantages of opening a dataset. Process transparency represents how information describes a process and its value objective is complete and available for the group of participants involved in the quasi-experiment case study (see Appendix B Part III).

We first tested if the variable was normally disturbed. Table 7-3 shows the normality test result whereby the significant number of the Shapiro-Wilk standard is 0,988, which means that the data is normally distributed (significant > 0,05).

Table 7-3 Tests of normality for the transparency process

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardised Residual for Transparency	0,045	147	0,200*	0,997	147	0,988

Next, we continue our test to create a descriptive analysis. Table 7-4 presents the descriptive statistics using MANOVA analysis in terms of the transparency process of the DSOD prototype. The result shows that BbN is most transparent for academia (mean = 97,00), followed by the governments and community (mean = 93,72 and 87,44 respectively). FMCDM contributes most to the academia group (mean = 89,73), followed by the government and community (mean = 88,89 and 83,38, respectively). In comparison, the DTA method was the most transparent process for the community group (mean = 93,25), followed by the government and academia (mean = 90,83 and 82,40, respectively). In sum, the best method is BbN for transparency.

Table 7-4 Descriptive statistic of the transparency process

Method		Mean	Std. Deviation	N
Bayesian-belief Networks	Government	93,72	5,603	18
	Academia	97,00	2,236	15
	Community	87,44	2,366	16
	Total	92,67	5,452	49
Fuzzy Multi-criteria Decision-making	Government	88,89	4,588	18
	Academia	89,73	4,148	15
	Community	83,38	2,986	16
	Total	87,35	4,816	49
Decision Tree Analysis	Government	90,83	2,065	18
	Academia	82,40	2,414	15
	Community	93,25	4,359	16
	Total	89,04	5,481	49
Total	Government	91,15	4,712	54
	Academia	89,71	6,727	45
	Community	88,02	5,241	48
	Total	89,69	5,678	147

Furthermore, Table 7-5 presents the Multivariate Tests (MANOVA) for the transparency independent variable. In MANOVA, there are four tests in each row: Pillai's Trace, Wilks' Lambda, Hotelling's Trace and Roy's Largest Root (Nath & Pavur, 1985). These different multivariate statistics test the statistical significance of the independent variables (Elliot, 2006; Nath & Pavur, 1985). The difference between the four-measure is how they combine the dependent variables (BbN, FMCDM, DTA) to examine the amount of variance in the data (Todorov & Filzmoser, 2010).

For this study, Wilks' Lambda multivariate measure is important. We looked for the independent variables' variance accounted for in the independent variables (BbN, FMCDM, and DTA) (transparency, accuracy, etc., perceived ease of use, and usefulness). The smaller the Wilks' Lambda value, the larger the difference between the groups being analysed (Elliot, 2006).

The Wilks' Lambda scale ranges from 0 to 1, where 0 means total discrimination (a significant difference), and 1 means no discrimination (no significant difference) between the dependent variables (Todorov & Filzmoser, 2010). Our statistical analysis using SPSS presented in Table 7-5 shows that a significant score ($\alpha = 0.108 > 0.05$). The F-test score is 1.310 is greater than the critical value (0.670), which is not significantly different from the means score of the dependent variables. This result interprets that there is no significant difference between the methods used (BbN, FMCDM, and DTA) in terms of the transparency process of the DSOD. In Table 7-5, the intercept part shows the constant mean value of every column (Y) in the variable tests through the all raw (X) equal to 0. In this case, we only focused on the Transparency effect.

Table 7-5 Multivariate Tests Result for the Transparency Variable

Multivariate Tests[a]						
Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.905	590.263[b]	2.000	124.000	.000
	Wilks' Lambda	.095	590.263[b]	2.000	124.000	.000
	Hotelling's Trace	9.520	590.263[b]	2.000	124.000	.000
	Roy's Largest Root	9.520	590.263[b]	2.000	124.000	.000
Transparency	Pillai's Trace	.361	1.311	42.000	250.000	.108
	Wilks' Lambda	.670	1.310[b]	42.000	248.000	.108
	Hotelling's Trace	.447	1.309	42.000	246.000	.109
	Roy's Largest Root	.287	1.707[c]	21.000	125.000	.038
a. Design: Intercept + Transparency						
b. Exact statistic						
c. The statistic is an upper bound on F that yields a lower bound on the significance level.						

7.2.3.2 Accuracy of the DSOD

A fundamental requirement of a good decision-making support system is that they result in accurate outcomes (Peignot, Peneranda, Amabile, & Marcel, 2013). Accuracy refers to the degree of compliance with the standard measurement, which reaches the actual measurement and is right on target. Accuracy measures the accuracy and similarity of results at the same time by comparing them to absolute values. In our case, the accuracy is based on subjective measures, as the participants estimated the accuracy of each method.

Table 7-6 shows the normality test result whereby the Significant number of the Shapiro-Wilk standard is 0,399, which means that the data is normally distributed (significant > 0,05).

Table 7-6 Tests of normality for the accuracy process

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardised Residual for Transparency	0,044	147	0,200*	0,990	147	0,399

Subsequently, we conducted a descriptive analysis. Table 7-7 presents the descriptive statistics. The result indicates that BbN is the most accurate rated by academia (mean = 98,80), followed by the governments and community (mean = 93,72 and 87,44 respectively). FMCDM is the most accurate to the academia group (mean = 93,20), followed by the community and government (mean = 91,38 and 88,89, respectively). In the comparison table, the DTA method was the most accurate for the government group (mean = 95,67), followed by the community and academia (mean = 93,25 and 82,40, respectively). In total, BbN method was rated as being the most accurate method.

Table 7-7 Descriptive statistic of the accuracy process

Method		Mean	Std. Deviation	N
Bayesian-belief Networks	Government	93,72	5,603	18
	Academia	98,80	1,424	15
	Community	87,44	2,366	16
	Total	93,22	5,868	49
Fuzzy Multi-criteria Decision-making	Government	88,89	4,588	18
	Academia	93,20	5,634	15
	Community	91,38	8,016	16
	Total	91,02	6,326	49
Decision Tree Analysis	Government	95,67	4,459	18
	Academia	82,40	2,414	15
	Community	93,25	4,359	16
	Total	90,82	6,900	49
Total	Government	92,76	5,610	54
	Academia	91,47	7,745	45
	Community	90,69	5,861	48
	Total	91,69	6,429	147

Table 7-8 shows the Multivariate Tests (MANOVA) for the accuracy of the result in using the DSOD. Based on our statistical analysis presented in Table 7-8 shows that the F-test score is 1.310, which is greater than the critical value (0.810) with a score $\alpha = 0.791 > 0.05$. This result indicates that there is no significant difference between the methods used (BbN, FMCDM, and DTA) in terms of the accuracy result of the DSOD.

Table 7-8 Multivariate Tests Result for the Accuracy Variable

Multivariate Tests[a]						
Effect		**Value**	**F**	**Hypothesis df**	**Error df**	**Sig.**
Intercept	Pillai's Trace	.863	390.436[b]	2.000	124.000	.000
	Wilks' Lambda	.137	390.436[b]	2.000	124.000	.000
	Hotelling's Trace	6.297	390.436[b]	2.000	124.000	.000
	Roy's Largest Root	6.297	390.436[b]	2.000	124.000	.000
Accuracy	Pillai's Trace	.239	.810	42.000	250.000	.792
	Wilks' Lambda	.773	.810[b]	42.000	248.000	.791
	Hotelling's Trace	.277	.811	42.000	246.000	.790
	Roy's Largest Root	.191	1.139[c]	21.000	125.000	.318
a. Design: Intercept + Accuracy						
b. Exact statistic						
c. The statistic is an upper bound on F that yields a lower bound on the significance level.						

7.2.3.3 Perceived Ease of Use the DSOD

In this study, we defined the perceived ease of use as the degree to which a person believes that the DSOD prototype is easy to understand. The perception of the ease of use of technology refers to a measure by which a person believes and knowledge that the technology can be easily understood and used (Hoffmann, 2016). The perceived ease of use can reduce a person's effort both time and effort to study a system or technology because individuals believe that the system or technology is easy to understand. The intensity of use and interaction between the user (user) and the system can also indicate the ease of use. The more frequently used systems indicate that they are more familiar, easier to operate, and easier to use by users.

Table 7-9 shows the normality test result whereby the Significant number of the Shapiro-Wilk standard is 0,894, which means that the data is normally distributed (significant > 0,05).

Table 7-9 Tests of normality for the ease of use

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardised Residual for Transparency	0,057	147	0,200*	0,995	147	0,894

Table 7-10 presents the descriptive statistics using multivariate ANOVA analysis. The result indicates that BbN is the easiest of use for the community group (mean = 97,63), followed by the governments and academia (mean = 93,72 and 87,44 respectively). FMCDM is the easiest to use for the group from academia (mean = 90,47), followed by the government and community (mean = 88,56 and 84,06, respectively). The DTA method was the easiest to use in the comparison table for the community group (mean = 95,56), followed by the government and academia (mean = 93,94 and 83,40, respectively).

Table 7-10 Descriptive statistic ease of use

Method		Mean	Std. Deviation	N
Bayesian-belief Networks	Government	94,00	5,821	18
	Academia	88,40	2,063	15
	Community	97,63	2,446	16
	Total	93,47	5,386	49
Fuzzy Multi-criteria Decision-making	Government	88,56	4,853	18
	Academia	90,47	3,833	15
	Community	84,06	3,130	16
	Total	87,67	4,771	49
Decision Tree Analysis	Government	93,94	4,137	18
	Academia	83,40	2,746	15
	Community	95,56	3,425	16
	Total	91,24	6,333	49
Total	Government	92,17	5,528	54
	Academia	87,42	4,175	45

Method		Mean	Std. Deviation	N
	Community	92,42	6,719	48
	Total	90,80	5,995	147

Table 7-11 shows the Multivariate Tests (MANOVA) for the perceived ease of use of the DSOD. The MANOVA statistical analysis shows that the F-test score is 1.143, which is higher than the critical value (0.702) with a score of α = 0.265, which is also higher than 0.05. This result shows that there is no significant difference between the methods used (BbN, FMCDM, and DTA) in terms of the perceived ease of use of the DSOD prototype.

Table 7-11 Multivariate Tests Result for the perceived ease of use Variable

Multivariate Tests[a]						
Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.894	524.473[b]	2.000	124.000	.000
	Wilks' Lambda	.106	524.473[b]	2.000	124.000	.000
	Hotelling's Trace	8.459	524.473[b]	2.000	124.000	.000
	Roy's Largest Root	8.459	524.473[b]	2.000	124.000	.000
Percieved ease of use	Pillai's Trace	.324	1.151	42.000	250.000	.255
	Wilks' Lambda	.702	1.143[b]	42.000	248.000	.265
	Hotelling's Trace	.387	1.134	42.000	246.000	.276
	Roy's Largest Root	.213	1.268[c]	21.000	125.000	.210
a. Design: Intercept + Easiness						
b. Exact statistic						
c. The statistic is an upper bound on F that yields a lower bound on the significance level.						

7.2.3.4 Usefulness of the DSOD

Perceived usefulness is the extent to which an individual believes that using a certain system will improve his performance (Meijer & Thaens, 2009). The definition of perceived usefulness is that users' subjective probability of using the application system can increase their expectations (Bonczek et al., 1980; Meijer & Thaens, 2009). Furthermore, perceived usefulness reflects the subjective probability that users will use the new decision-making support, whether it will benefit themselves or their organisation. Cognitive factors also play an important role. The factors include the relevance of DSDO to the user and the individual's perception.

Table 7-12 shows the normality test result whereby the significant number of the Shapiro-Wilk standard is 0,088, which means that the data is normally distributed (significant > 0,05).

Table 7-12 Tests of normality of usefulness

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardised Residual for Transparency	0,082	147	0,016	0,984	147	0,088

Table 7-13 presents the descriptive statistics using multivariate ANOVA analysis for perceived usefulness. The result suggests that BbN is the most useful for academia (mean = 95,73), followed by the governments and academia (mean = 85,94 and 85,50 respectively). FMCDM is the most useful for the community group (mean = 84,94), followed by the academia and government (mean = 82,53 and 78,33, respectively). The DTA method was the easiest to use in the comparison table for the government group (mean = 95,89), followed by the academia and community (mean = 86,40 and 81,19, respectively). In total, DTA is perceived as the most useful method.

Table 7-13 Descriptive statistic usefulness

Method		Mean	Std. Deviation	N
Bayesian-belief Networks	Government	85,94	6,725	18
	Academia	95,73	3,173	15
	Community	85,50	6,398	16
	Total	88,80	7,311	49
Fuzzy Multi-criteria Decision-making	Government	78,33	5,466	18
	Academia	82,53	3,701	15
	Community	84,94	9,169	16
	Total	81,78	6,986	49

Method		Mean	Std. Deviation	N
Decision Tree Analysis	Government	95,89	5,989	18
	Academia	86,40	4,085	15
	Community	81,19	7,609	16
	Total	88,18	8,674	49
Total	Government	86,72	9,394	54
	Academia	88,22	6,653	45
	Community	83,88	7,881	48
	Total	86,25	8,277	147

Table 7-14 shows the Multivariate Tests (MANOVA) for the usefulness of the DSOD. Based on the Wilks' Lambda, the F score is 0.878 with a score α = 0.717 > 0.05, which means that there is no significant difference between the methods used in terms of the usefulness of the DSOD.

Table 7-14 Multivariate Tests Result for the usefulness Variable

Multivariate Tests[a]						
Effect		**Value**	**F**	**Hypothesis df**	**Error df**	**Sig.**
Intercept	Pillai's Trace	.905	552.921[b]	2.000	116.000	.000
	Wilks' Lambda	.095	552.921[b]	2.000	116.000	.000
	Hotelling's Trace	9.533	552.921[b]	2.000	116.000	.000
	Roy's Largest Root	9.533	552.921[b]	2.000	116.000	.000
Usefulness	Pillai's Trace	.358	.881	58.000	234.000	.713
	Wilks' Lambda	.672	.878[b]	58.000	232.000	.717
	Hotelling's Trace	.442	.876	58.000	230.000	.721
	Roy's Largest Root	.277	1.119[c]	29.000	117.000	.328
a. Design: Intercept + Usefulness						
b. Exact statistic						
c. The statistic is an upper bound on F that yields a lower bound on the significance level.						

In the e-procurement case study, our quasi-experiment resulted in several important findings. BbN appears to be the best method associated with the transparency and accuracy of the DSOD process. In terms of ease of use of the DSOD prototype, BbN, and DTA are the two most feasible methods to help stakeholders analyse the advantages and disadvantages of the opening datasets. For the stakeholders who consider the time efficiency in the decision-making process, the DTA is the best method to be employed.

7.3 Quasi-experiment 2: Deciding on opening medical records dataset

The second empirical setting of this research is in the medical records case study. The medical records dataset contains information about the history of patients and assessment or evaluation records of medical treatment. These medical records are an essential and crucial part of patient care planning and coordination for further medical treatment and ensuring the continuity of the clinic or hospital services.

In a situation like Indonesia's government, the opening of medical records makes it easier for the department of health and related stakeholders to monitor the current issues of the disease pandemic. Nevertheless, the number of medical records datasets published to the public may contain several disadvantages. The potential penalties encounter from the medical records dataset's opening are the possibility of personal privacy violation, the opening of inaccurate patient information, and contradicting or against the law.

Table 7-15 characteristics of the group participants to conduct the experiment

	Quasi-experiment 2A	Quasi-experiment 2B	Quasi-experiment 2C
Date	30 July 2019	1 August 2019	9 August 2019
Duration	145 minutes	115 minutes	125 minutes
Number of participants	15	12	10
Type of group participants	Government institution	Academia	Community
Location	Indonesia	Indonesia	Indonesia
Step 1. Introduction	✔	✔	✔
Step 2. Pre-test	✔	✔	✔
Step 3. Performing scenarios	✔	✔	✔
Step 4. Post-test	✔	✔	✔
Step 5. Discussion	✔	✔	✔

Table 7-15 shows the group's characteristics in the medical records case study. We used pre-test and post-test approaches to perform the quasi-experiments. The participants from the three different groups were selected non-randomly. These beta testings were carried out from 30 July to 9 August 2020 in Indonesia. The quasi-experiments took 145 minutes in total for participants from the local government institution. Government participants are working for the public information disclosure, department of public health, and social services department. The quasi-experiment at the academia group took 115 minutes and involved 15 participants from academia, e.g., associate professor and PhD candidates in the open data research area, lecturers, senior researchers, and master’s students from Telkom University and Bina Darma University, Indonesia. Furthermore, 16 participants in the community group included non-government organisations, such as Indonesia Corruption Watch (ICW) and Open Data Labs Jakarta. The participants from the communities included data analytics for analysing the potential misuse of the e-health patient dataset.

7.3.1 Demographics of the participants

This section discusses the characteristics and differences between the quasi-experiment participants. This study used several sub-questions to characterise the participants' demographics, including gender, age, educational level, organisation type, and current job function of the participants. In total, there were 37 participants involved in this quasi-experiment case study. In all three quasi-experiments, the majority of the participants were male (54%). The age distributions reveal that most of the participants fall from 35 to 44 years old (32%), as shown in Figures 7-6.

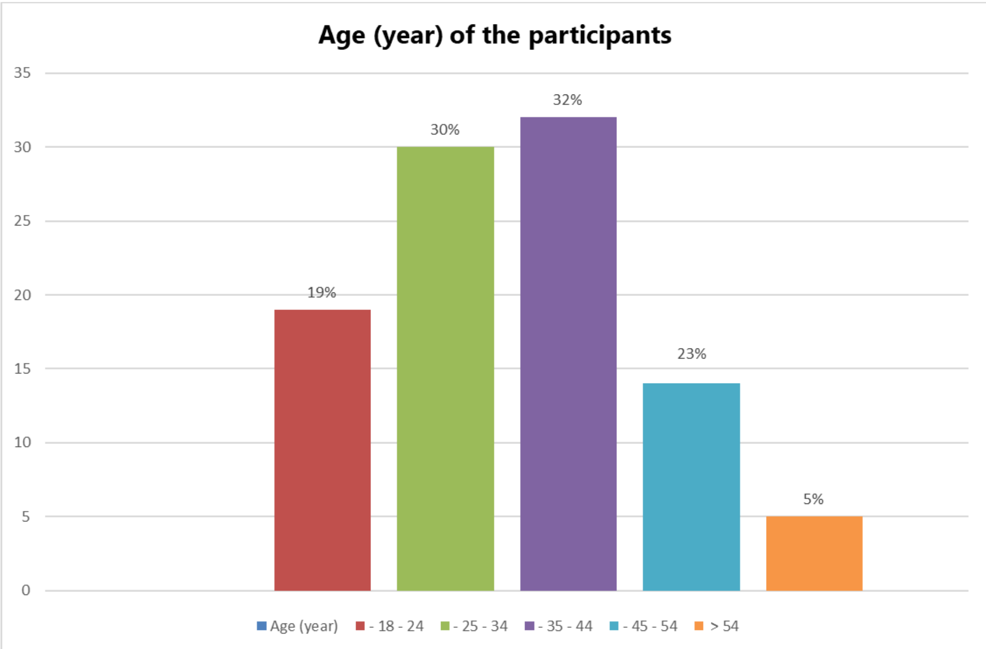


Figure 7-6 Distribution of the age of the participants

From the education background, the survey shows that most of the participants had a master graduate level (62%), followed by the bachelor and doctorate level (22% and 16%, respectively). This result indicates that the participants are highly educated.

Figure 7-97 shows the type of organisations of the participants. From the organisational types, our survey indicates that most of the contributors were from a local government institution (35%), followed by the participants from university or academia and community-based organisation (33% and 16%, correspondingly). Simultaneously, other participants came from business or private sectors (11%) and governments or ministries (5%). From this figure, we found that the participants' characteristic based on their home organisations is significantly different distributed.

Figure 7-8 shows the job function of the participants. Regarding the job description of the participants, there are nine job positions in their organisations. Most participants are responsible for analysing and investigating datasets from the OGD portal (33%). The participants' second job function is referred to as being responsible for deciding the public policies, followed by providing services directly to the public (16% and 11%, separately). Other job roles can be classified as administrative support (8%), technical services (8%), coordinate one or more programs (8%), and technical specialist (8%).

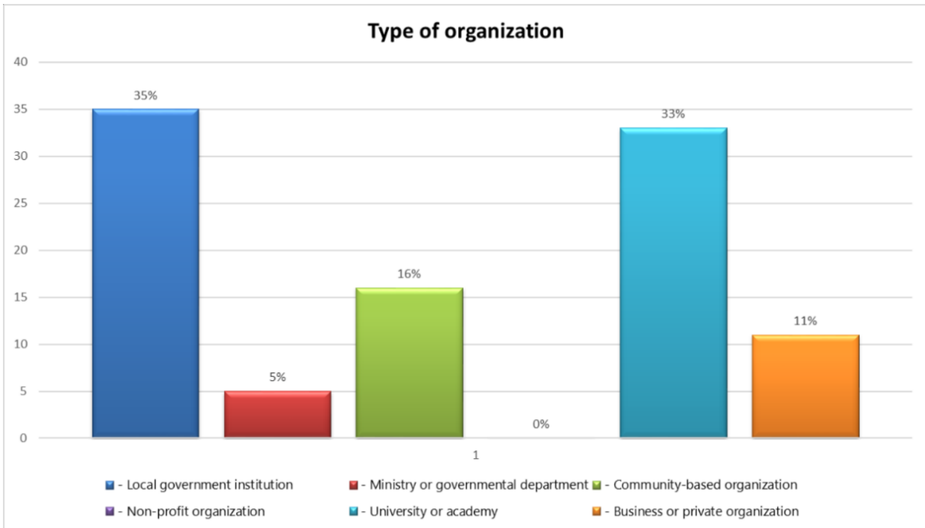


Figure 7-7 Type of organisation of the quasi-experiment participants

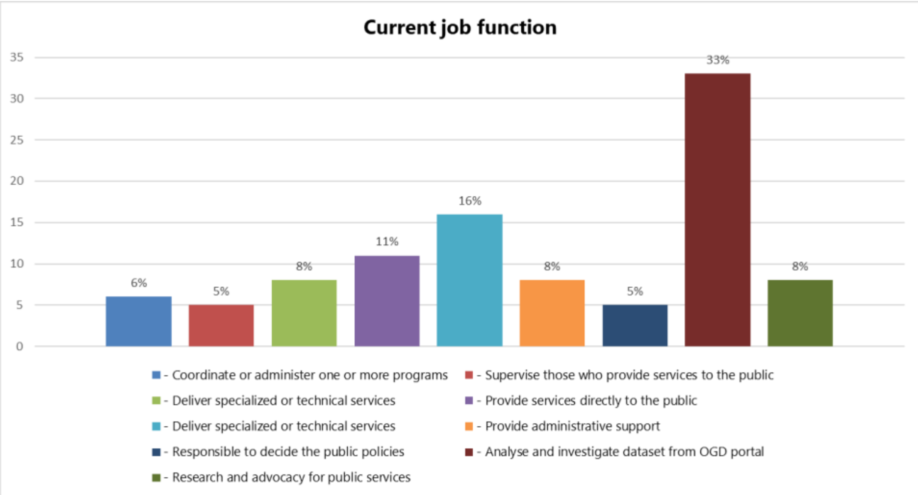


Figure 7-8 Current job function of the quasi-experiment participants

7.3.2 Experience in using open government data

This section elaborates on the characteristics and the differences between the quasi-experiment participants from the perspective of experience using open government data. This study used four characteristics to represent the participants' experience and expertise in using OGD datasets. First, from the year of experience in using the OGD dataset, most participants used the open datasets were between 6 to 10 years (51%), followed by the between 1 to 5 years (30%), and some even used open datasets already between 11 to 15 years (16%).

Figure 7-9 shows the frequency of the participants in using open government datasets. Our study found that most of the participants used the OGD datasets daily or multiple times per day is more than half participants (62%), followed by the monthly or a few times per month to use OGD datasets (16%). Furthermore, 16% of the participants used the OGD dataset for their weekly job activities. About 6% of the participants were rarely using the OGD dataset.

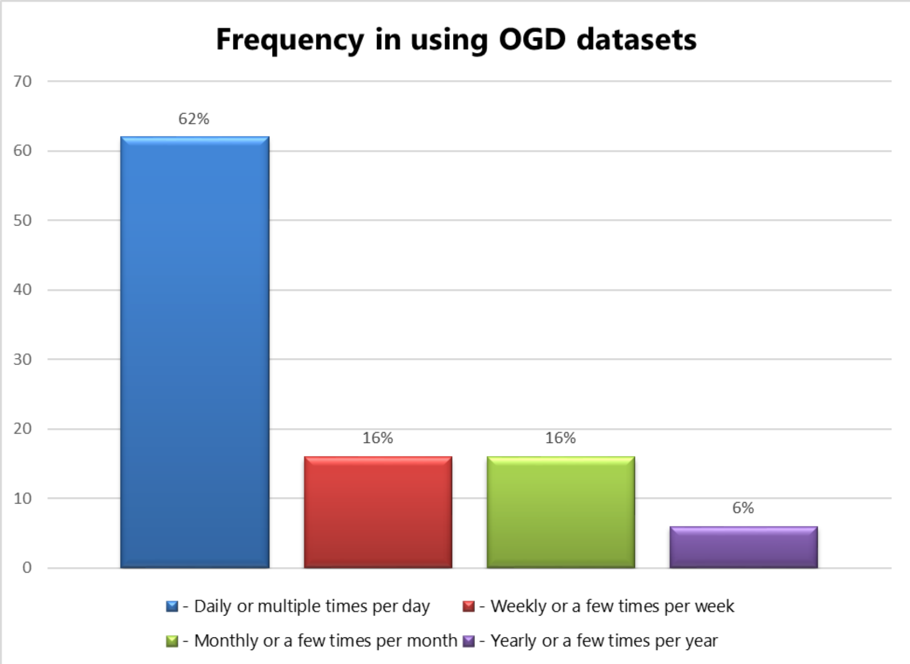


Figure 7-9 Frequency in using OGD of the quasi-experiment participants

7.3.3 Measurement of the quasi-experiment

In this research, we used four main variables to measure the effects of the developed prototype. The use of three different methods (BbN, FMCDM, and DTA) was evaluated because each method performs to have different benefits and objectives. A quasi-experiment setting was evaluated and analysed using multivariate ANOVA analysis to test the different participant groups using three methods (BbN, FMCDM, and DTA).

7.3.3.1 Transparency of the DSOD

Table 7-16 shows the normality test result whereby the Significant number of the Shapiro-Wilk standard is 0.482, which means that the data is normally distributed (significant > 0,05).

Table 7-16 Tests of normality for the transparency process

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardised Residual for Transparency	0,058	111	0,200*	0,989	111	0,482

After that, we continue our test to create a descriptive analysis. Table 7-17 presents the descriptive statistics using multivariate MANOVA analysis in terms of the transparency process of the DSOD prototype. The result indicates that BbN is the most transparent for the government (mean = 93,60), followed by the academia and community (mean = 80,50 and 77,20 respectively). FMCDM is the most transparent to the academia group (mean = 95,83), followed by the government and community (mean = 72,47 and 55,80, respectively). In comparison, the DTA method was the most transparent process for the government group (mean = 99,27), followed by the community and academia (mean = 84,10 and 80,92, respectively). In total, the best method for the transparency process of the DSOD prototype is DTA.

Table 7-17 Descriptive statistic of the transparency

Method		Mean	Std. Deviation	N
Bayesian-belief Networks	Government	93,60	5,396	15
	Academia	80,50	3,873	12
	Community	77,20	8,217	10
	Total	84,92	9,340	37
Fuzzy Multi-criteria Decision-making	Government	72,47	8,202	15
	Academia	95,83	6,699	12
	Community	55,80	5,391	10
	Total	75,54	17,222	37
Decision Tree Analysis	Government	99,27	1,163	15
	Academia	80,92	2,575	12
	Community	84,10	6,806	10
	Total	89,22	9,298	37

Total	Government	88,44	12,927	45
	Academia	85,75	8,557	36
	Community	72,37	13,947	30
	Total	83,23	13,668	111

Table 7-18 shows the Multivariate Tests (MANOVA) for the transparency of the DSOD in the case study of the e-Health patient dataset. Similar to the first quasi-experiment, we also use Wilks' Lambda score to find the amount of variance accounted for in the dependent variables (BbN, FMCDM, and DTA) by the independent variables (transparency, accuracy, perceived ease of use, and usefulness). Based on the Wilks' Lambda, the F score is 1.401 with a value $\alpha = 0.059 > 0.05$), indicating there is no significant difference between the methods used in terms of the transparency process of the DSOD.

Table 7-18 Multivariate Tests Result for the transparency Variable

Multivariate Tests[a]						
Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.866	270.720[b]	2.000	84.000	.000
	Wilks' Lambda	.134	270.720[b]	2.000	84.000	.000
	Hotelling's Trace	6.446	270.720[b]	2.000	84.000	.000
	Roy's Largest Root	6.446	270.720[b]	2.000	84.000	.000
Transparency	Pillai's Trace	.549	1.288	50.000	170.000	.120
	Wilks' Lambda	.498	1.401[b]	50.000	168.000	.059
	Hotelling's Trace	.912	1.514	50.000	166.000	.028
	Roy's Largest Root	.792	2.691[c]	25.000	85.000	.000
a. Design: Intercept + Transparency						
b. Exact statistic						
c. The statistic is an upper bound on F that yields a lower bound on the significance level.						

7.3.3.2 Accuracy of the DSOD

Table 7-19 shows the normality test result whereby the Significant number of the Shapiro-Wilk standard is 0,512, which means that the data is normally distributed (significant > 0,05).

Table 7-19 Tests of normality for the accuracy process of the DSOD prototype

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardised Residual for Transparency	0,067	111	0,200*	0,989	111	0,512

Table 7-20 presents the descriptive statistics using multivariate ANOVA analysis in terms of the accurate process of the DSOD prototype. The result indicates that BbN is the most accurate for academia (mean = 98,42), followed by the community and government (mean = 80,50 and 76,13, respectively). FMCDM is the most to the academia group (mean = 96,92), followed by the community and government (mean = 89,40 and 82,47, respectively). In the comparison table, the DTA method was the most transparent process for the community group (mean = 92,20), followed by the community and academia (mean = 78,80 and 70,75, respectively). In total, the most accurate process of the DSOD prototype is BbN.

Table 7-20 Descriptive statistic of the accuracy

Method		Mean	Std. Deviation	N
Bayesian-belief Networks	Government	76,13	4,998	15
	Academia	98,42	2,466	12
	Community	80,50	4,601	10
	Total	84,54	10,725	37
Fuzzy Multi-criteria Decision-making	Government	82,47	7,367	15
	Academia	96,92	6,762	12
	Community	89,40	7,691	10

	Total	89,03	9,412	37
Decision Tree Analysis	Government	78,80	10,758	15
	Academia	70,75	8,884	12
	Community	92,20	6,730	10
	Total	79,81	12,283	37
Total	Government	79,13	8,303	45
	Academia	88,69	14,390	36
	Community	87,37	8,045	30
	Total	84,46	11,415	111

Table 7-21 indicates the Multivariate Tests (MANOVA) for the accuracy of the DSOD. Based on the Wilks' Lambda, the F score is 1.525 with an α = 0.019 < 0.05, which means that there is a significant difference between the methods used in terms of the accuracy result of the DSOD.

Table 7-21 Multivariate Tests Result for the accuracy variable

Multivariate Tests[a]						
Effect		**Value**	**F**	**Hypothesis df**	**Error df**	**Sig.**
Intercept	Pillai's Trace	.908	381.726[b]	2.000	77.000	.000
	Wilks' Lambda	.092	381.726[b]	2.000	77.000	.000
	Hotelling's Trace	9.915	381.726[b]	2.000	77.000	.000
	Roy's Largest Root	9.915	381.726[b]	2.000	77.000	.000
Accuracy	Pillai's Trace	.775	1.541	64.000	156.000	.016
	Wilks' Lambda	.375	1.525[b]	64.000	154.000	.019
	Hotelling's Trace	1.272	1.510	64.000	152.000	.021
	Roy's Largest Root	.712	1.737[c]	32.000	78.000	.025
a. Design: Intercept + Accuracy						
b. Exact statistic						
c. The statistic is an upper bound on F that yields a lower bound on the significance level.						

7.3.3.3 Perceived Ease of use

Table 7-22 shows the normality test result whereby the Significant number of the Shapiro-Wilk standard is 0,437, which means that the data is normally distributed (significant > 0,05).

Table 7-22 Tests of normality for the ease of use

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
Standardised Residual for Transparency	0,063	111	0,200*	0,988	111	0,437

Table 7-23 presents the descriptive statistics using multivariate ANOVA analysis in terms of the transparency process of the DSOD prototype. The results show that BbN is the easiest for the academia group (mean = 94,75), followed by the community and academia (mean = 79,30 and 63,73 respectively). FMCDM is the easiest of use to the academia group (mean = 92,17), followed by the community and government (mean = 85,10 and 77,33, separately). In the comparison table, the DTA method was easiest for the government group (mean = 95,47), followed by the community and academia (mean = 88,00 and 70,75, respectively). The easiest of use process of the DSOD prototype is DTA.

Table 7-23 Descriptive statistic ease of use

Method		Mean	Std. Deviation	N
Bayesian-belief Networks	Government	63,73	2,576	15
	Academia	94,75	5,101	12
	Community	79,30	3,592	10
	Total	78,00	13,876	37
Fuzzy Multi-criteria Decision-making	Government	77,33	5,150	15
	Academia	92,17	6,308	12
	Community	85,10	10,027	10
	Total	84,24	9,415	37

Decision Tree Analysis	Government	95,47	2,850	15
	Academia	70,75	8,884	12
	Community	88,00	5,121	10
	Total	85,43	12,226	37
Total	Government	78,84	13,636	45
	Academia	85,89	12,826	36
	Community	84,13	7,542	30
	Total	82,56	12,315	111

Table 7-24 indicates the Multivariate Tests (MANOVA) for the perceived ease of use of the DSOD. Based on the Wilks' Lambda, the F score is 0.286 with a significant score α = 0.004 < 0.05. This suggests a significant difference between the methods used in terms of the perceived ease of use of the DSOD.

Table 7-24 Multivariate Tests Result for the perceived ease of use variable

Multivariate Tests[a]						
Effect		**Value**	**F**	**Hypothesis df**	**Error df**	**Sig.**
Intercept	Pillai's Trace	.938	544.543[b]	2.000	72.000	.000
	Wilks' Lambda	.062	544.543[b]	2.000	72.000	.000
	Hotelling's Trace	15.126	544.543[b]	2.000	72.000	.000
	Roy's Largest Root	15.126	544.543[b]	2.000	72.000	.000
Easiness	Pillai's Trace	.920	1.681	74.000	146.000	.004
	Wilks' Lambda	.286	1.696[b]	74.000	144.000	.004
	Hotelling's Trace	1.782	1.710	74.000	142.000	.003
	Roy's Largest Root	1.163	2.294[c]	37.000	73.000	.001
a. Design: Intercept + Easiness						
b. Exact statistic						
c. The statistic is an upper bound on F that yields a lower bound on the significance level.						

7.3.3.4 Usefulness of the DSOD

Table 7-25 shows the normality test result whereby the Significant number of the Shapiro-Wilk standard is 0,176, which means that the data is normally distributed (significant > 0,05).

Table 7-25 Tests of normality of usefulness

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardised Residual for Transparency	0,098	111	0,010	0,983	111	0,176

Next, we continue our test to create a descriptive analysis. Table 7-26 presents the descriptive statistics using multivariate ANOVA analysis in terms of the DSOD prototype's usefulness. The result indicates that BbN is the most useful for the government group (mean = 99,47), followed by the community and academia (mean = 85,94 and 85,50 respectively). FMCDM is the most useful for the government group (mean = 93,60), followed by the academia and community (mean = 80,50 and 77,20, respectively). The DTA method was the easiest to use in the comparison table for the government group (mean = 95,53), followed by the academia and community (mean = 85,92 and 80,20, respectively). In total, the best transparent process of the DSOD prototype is BbN.

Table 7-26 Descriptive statistic usefulness

Method		Mean	Std. Deviation	N
Bayesian-belief Networks	Government	99,47	0,990	15
	Academia	81,17	2,368	12
	Community	81,30	2,627	10
	Total	88,62	9,287	37
Fuzzy Multi-criteria Decision-making	Government	93,60	5,396	15
	Academia	80,50	3,873	12

	Community	77,20	8,217	10
	Total	84,92	9,340	37
Decision Tree Analysis	Government	95,53	5,383	15
	Academia	85,92	7,585	12
	Community	80,20	8,337	10
	Total	88,27	9,389	37
Total	Government	96,20	4,989	45
	Academia	82,53	5,526	36
	Community	79,57	6,912	30
	Total	87,27	9,404	111

Table 7-27 indicates the Multivariate Tests (MANOVA) for the usefulness of the DSOD. Based on the Wilks' Lambda, the F score is 0.248 with a value α = 0.000 < 0.05. This suggests that there is a significant difference between the methods used regarding the usefulness of the DSOD.

Table 7-27 Multivariate Tests Result for the usefulness variable

Multivariate Tests[a]						
Effect		**Value**	**F**	**Hypothesis df**	**Error df**	**Sig.**
Intercept	Pillai's Trace	.932	581.712[b]	2.000	85.000	.000
	Wilks' Lambda	.068	581.712[b]	2.000	85.000	.000
	Hotelling's Trace	13.687	581.712[b]	2.000	85.000	.000
	Roy's Largest Root	13.687	581.712[b]	2.000	85.000	.000
Usefulness	Pillai's Trace	.946	3.213	48.000	172.000	.000
	Wilks' Lambda	.248	3.564[b]	48.000	170.000	.000
	Hotelling's Trace	2.244	3.927	48.000	168.000	.000
	Roy's Largest Root	1.813	6.497[c]	24.000	86.000	.000
a. Design: Intercept + Usefulness						
b. Exact statistic						
c. The statistic is an upper bound on F that yields a lower bound on the significance level.						

In summary, in the second quasi-experiment, DTA is the best method associated with the transparency, ease of use, and usefulness of the DSOD. In terms of ease of use of the DSOD prototype, DTA is the most feasible method to help stakeholders analyse the advantages and disadvantages of the opening datasets. For the stakeholders who consider the accuracy in the decision-making process, the Bayesian-belief Network is the best method to be employed.

7.4 Comparative analysis of two quasi-experiments

Using various criteria for evaluating the DSOD, it was found that each method had its unique pros and cons. Table 7-28 compares the methods for their level of transparency, accuracy, ease of use, and usefulness for the two case studies. The mean score in Table 7-28 indicates the average performance of the dependent variables (BbN, FMCDM, and DTA) dividing by the average score of the independent variables (transparency, accuracy, easiness, and usefulness).

Table 7-28 Overview of the comparison result between variable analysis

Case study	Variable	The best method	Mean score
e-procurement	Transparency	BbN	97,00
	Accuracy	BbN	98,80
	Easiness	BbN	97,63
	Usefulness	DTA	95,89
Medical records	Transparency	DTA	99,27
	Accuracy	BbN	98,42
	Easiness	DTA	95,47
	Usefulness	BbN	99,47

In the e-procurement case study, our quasi-experiment identified several important findings. First, in terms of the transparency process of the DSOD, BbN appears to be the best method to be employed in the decision-making process to analyse the datasets (mean score 97,00) compare to two other methods. In this case, the academic group believed that BbN could help them make the decision-making process more transparent. Second, related to the analysis's accuracy, BbN presented the best performance to guarantee the process's results are more precise (mean score 98,80). Our experiment shows that researchers, lecturers, and students believe that BbN can cover the accuracy of the decision-making process to analyse the datasets. Third, regarding the perceived ease of use of the DSOD, the BbN seems the best method in the e-Procurement case study. The community believes that the BbN presented the systematic steps and visual causal relationship diagram to understand better the decision-making process. Fourth, however, the differences in the tendency to use the usefulness methods, DTA appeared to be the best systematic method to analyse the e-procurement dataset (95,89). Based on our experiment, government institutions prefer to use the DTA to analyse the potential costs and benefits of disclosing data.

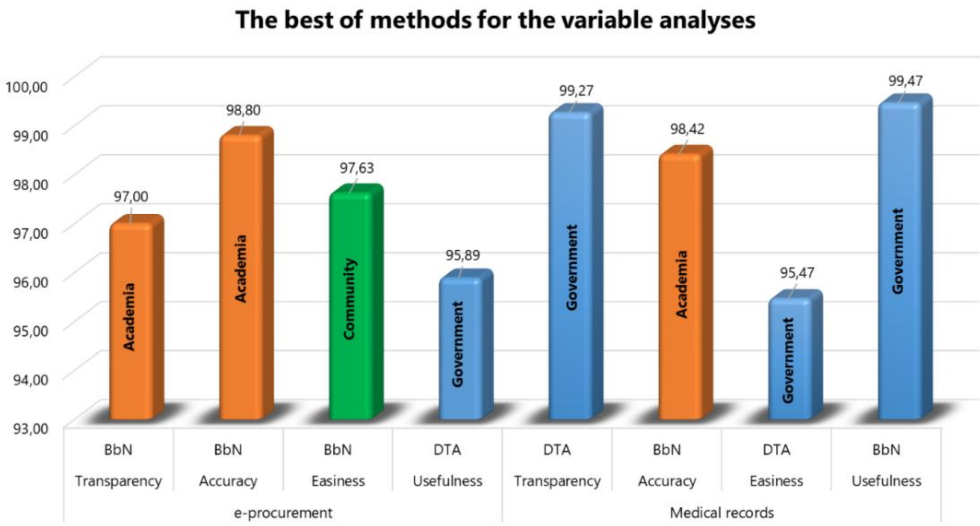


Figure 7-10 Comparison of the most feasibility of the methods

Furthermore, Figure 7-10 shows the comparison between the methods regarding the level of transparency, accuracy, perceived ease of use, and usefulness for the medical records case study. In terms of the transparency process of the DSOD, DTA appears to be the best method to be employed in the decision-making process to analyse the datasets (Score = 99,27). In this case, three different groups of participants, such as government organisations, were convinced that DTA would help decision-makers and other government employees and make the decision-making process more transparent. Second, related to the analysis's accuracy, BbN scored highest on accuracy (Score = 98,42). Third, regarding the ease of use of the DSOD, the DTA was successfully performed in the medical records case study (Score = 95,47). From the government institutions' perspective, the DTA showed the well-structured steps and its causal relationship diagram that better understand the decision-making process. Regarding the usefulness of the DSOD, BbN was selected systematically to analyse the potential advantages and disadvantages of opening medical record datasets (Score = 99,47). Our experiment shows that the government side participants were confident to employ BbN because of their relevance and benefits to their current works.

Table 7-29 Overview of the comparison result between the group of participants

Case study	Group	The best method	Mean score
e-Procurement	Government	DTA	95,89
	Academia	BbN	97,00
	Community	BbN	97,63
Medical records	Government	BbN	99,47
	Academia	BbN	98,42
	Community	DTA	92,20

Table 7-29 presents a comparison between the most feasible methods used in terms of the three different groups of participants in the two case studies. In the e-procurement case study, our experiment shows several essential findings. DTA is the most preferred method by government institutions (mean score = 95,89). The government actors like decision-makers, policy-makers, and administrative officers prefer to use the DTA since it could be useful for their current jobs. Second, BbN is viewed as the best alternative method for academia to analyse the potential advantages and disadvantages of opening datasets (mean score = 97,00). The participants from academia believed to use BbN as the best method because of its ability to show the transparency process in decision-making to analyse the e-procurement dataset. BbN is also the community participants' preferred method to assess the potential advantages and disadvantages of disclosing data mean score = 97,63). Community stakeholders, like professional data analytics, non-government organisations, and general users prefer BbN.

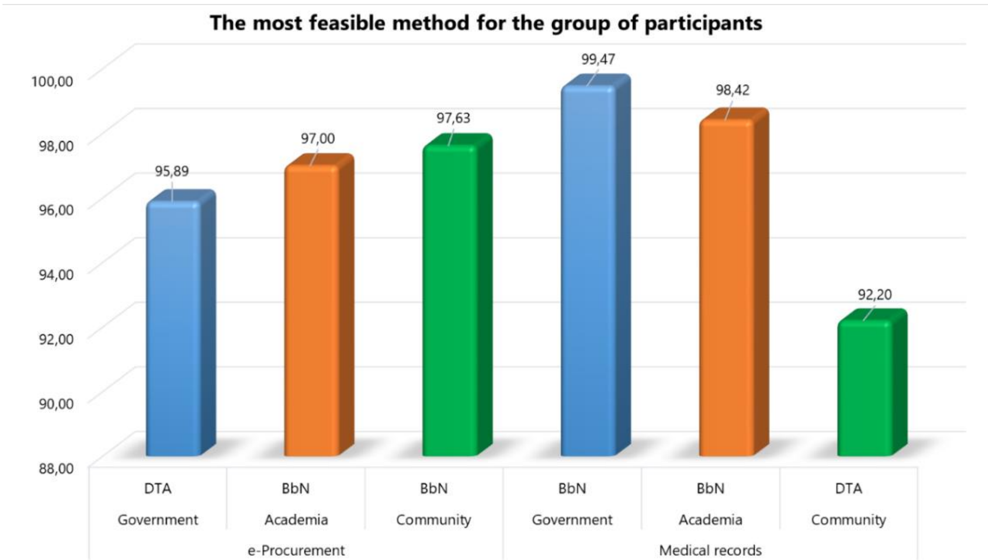


Figure 7-11 Comparison of the methods by groups of participants

Figure 7-11 compares the methods used in the medical records case study. First, our experimental case study shows that BbN is the most viable and preferred method for government institutions. Decision-makers and policy-makers are convinced that using this method could help them to analyse the advantages and disadvantages of disclosing data. Second, BbN is also preferred by higher educational institution participants, e.g., researchers, lecturers, and students, to assess the potential advantages and disadvantages of disclosing data. The participants from academia believe that BbN can perform systematic steps to provide more accurate results. Third, DTA is the community's preferred method to assess the potential advantages and disadvantages of releasing data. Community actors, such as professional data analytics and non-government organisations, preferred DTA since it provides more accuracy in the decision-making process to analyse the medical records dataset.

7.5 Conclusions

Based on our experiments, we conclude that the participants find that there is no single best method to help governments, academia, and communities analyse the potential advantages and disadvantages of opening the dataset. The selection of the methods depends on several factors. The factors might include the diversity of the organisational policy and their objectives, lack of human-skilled resources, different levels of formal education, and lack of experience in analysing the datasets. In the BbN method, the experts and decision-makers spent more time processing the evaluation compared to the other two methods. One reason is that in the BbN method, the time allocation to evaluate the selected dataset depended on the expert's input, which is subjective as their knowledge and experience are limited. Therefore, we suggest developing a repository when using historical data or expert's knowledge in further research. In addition, we found that the advantage of using BbN method is the result more accurate in constructing the uncertainty relationships. DTA is the most viable method to be employed in terms of the usefulness in the case study of e-procurement and medical records at the same time. Our experiment study indicated that FMCDM had fewer contributions to stakeholders' expectations to support the decision-making process to open data. However, based on our experimental case study, FMCDM could help academics analyse the advantages and disadvantages of the opening data regarding transparency and accuracy of the results in the decision-making process. Academia finds that FMCDM is a transparent method with detailed steps to weigh and use the Fuzzy linguistic matrix.

Our Multivariate Test (MANOVA) analysis indicates that in the case study of e-Procurement, all the independent variables (transparency, accuracy, perceived ease of use, and usefulness) have no significant difference in the three methods used. The MANOVA suggests that there were no significant differences among the methods for the DSOD. In contrast, in the second case study (eHealth dataset), we found that three independent variables, e.g. accuracy, perceived ease of use, and usefulness, have significant differences with regards to the methods used (BbN, FMCDM, and DTA). The methods perform differently on the level of accuracy, perceived ease of use and usefulness. Based on these results, we conclude that the different types of datasets with different levels of advantages and disadvantages, e.g., potential privacy violation, misinterpretation of the dataset, and other risk-averse categories, can influence the methods' significance.

Chapter 8
Conclusions and further research

This dissertation has developed a systematic approach to decision-making support to analyse the advantages and disadvantages of opening data. This chapter presents the overall conclusions of this research in Section 8.1 and reflects this with the existing literature. Next, the research questions are revisited in Section 8.2. We discuss why decision-making processes to open data are not trivial in Section 8.3. Thereafter research limitations are presented in Section 8.4, and further research recommendations in Section 8.5.

8.1 Overall conclusions

This research started with surveying the literature's pros and cons to disclose the public domain datasets. Politicians all too often focus only on the advantages of opening data like improving transparency and enhancing citizen engagement, whereas potential disadvantages are given less attention (Kassen, 2013; Josefin Lassinantti, Bergvall-Kåreborn, & Ståhlbröst, 2013). At the same time, public servants are reluctant to publish the datasets due to a focus on the potential disadvantages (Kucera & Chlapek, 2014; Zuiderwijk & Janssen, 2015) and the complexities with regards to the decision-making process (Martin et al., 2013; Veenstra & Broek, 2013). Furthermore, decision-makers seemed to keep and not to open the datasets because they face challenges related to privacy issues, legislation, and data protection (Czarnitzkia et al., 2020; Kulk & van Loenen, 2012; Scasa, 2014). The decision-makers might not know or are unaware of which dataset can be opened and alternatives beyond the binary decision to "open" or "close" a complete dataset. In addition, the involvement of diverse stakeholders looking at different aspects and interests complicates decision-making about the opening data. Some stakeholders prefer low-cost decision-making, while others prefer risk-averse decisions.

Although the literature provides several overviews of evaluating the potential disadvantages and other adverse effects in the open data domain, there is still no systematic approach to weigh these advantages and disadvantages. Previous research only emphasised a simple process for limited stakeholders rather than providing a more concrete decision-making process. There are only a few studies introduce method on how to deal with the possible disadvantages of opening data. (Brickell & Shmatikov, 2008) introduced a trade-off method to analyse sensitive attributes and benefits of a given dataset using an anonymisation algorithm. (Taneja, Kapil, & Singh, 2015) used a similar method using an anonymisation tool to re-identify the sensitive personal attributes in a selected dataset. These two example approaches focus on re-identifying the disadvantages of having personal data without considering potential benefits, the costs of opening or viewing other types of disadvantages like the opening of sensitive data and the opening of inappropriate data.

Our study developed a decision-making support model by weighing the advantages and disadvantages to open data. There was no systematic literature study to categorise the advantages and disadvantages of opening data. Therefore, a taxonomy of the possible benefits and drawbacks of disclosing data was developed in this research. The taxonomy enables us to evaluate the disadvantages and advantages of datasets structurally.

Furthermore, our literature study found that several methods could be used to analyse and weigh the open government data domain's advantages and disadvantages. However, it has not been clear what methods are most appropriate for use by the open government data decision-makers to analyse the potential advantages and disadvantages of disclosing data. Since there is no single best method found in the previous study, we decided to select three different methods (BbN, FMCDM, and DTA). By choosing different methods, we would able to compare the appropriateness of these methods.

BbN was selected to deal with the probability of an event occurring and its ability to take uncertainty into account. The BbN allows a combination of data with domain knowledge and facilitates learning about causal relationships between factors influencing the advantages and disadvantages of opening data. The BbN provides a principal approach of probability distribution rather than a point estimate. We postulate that the BbN method might be useful to organisations that have a limited number of experts to quantify the potential advantages and disadvantages of opening data.

FMCDM was selected to express different criteria to take into account various stakeholder interests and objectives. The FMCDM aims to manage problems in making decision alternatives to express these concerns with the computational approach. We postulate that the FMCDM might help academia and the community stakeholders with sufficient knowledge and experiences to analyse the advantages and disadvantages of opening data variables.

DTA was selected for its capability to consider all possible outcomes and a decision taken by the decision-makers and traces each path to a conclusion. The DTA method is relatively easy to apply both at the decision-makers and operational officer's level. The DTA can manage several factors of the likely costs and benefits of opening datasets. We presume that DTA might help data publishers of government organisations to estimate the possible consequence of the opening data.

Many decision-support works of literature focus on providing support for a single decision-maker. However, in this research, multiple stakeholders involved in decision-making and operating in several societal environments in the public domain were taken into account. DSOD supported different weights and criteria and was also selected. This study used a set of DSOD criteria, including transparency for public value, ease of use, expected results accuracy, and open data stakeholders' usefulness. This is likely the first study to compare different decision-making methods with each other. The analysis results show that a public value like the transparency of the outcomes comes at a price, as more efforts are needed, as shown by our case studies.

Our quasi-experiment shows that the participants prefer the BbN because it generates causal relationships between factors influencing the advantages and disadvantages of opening data. The BbN would help to calculate the probability and manipulate the degree of human belief. The FMCDM method was preferable by particular researchers since this method strongly depends on human knowledge and expertise. Our analyses also found important results, whereby the DTA is the most helpful and relevant to the current work for the decision-makers, policy-makers, civil servants, and administrative officers. The DTA provides simple steps to construct decision alternatives and can provide sufficient insight into open data decisions. The DTA creates a comprehensive analysis of the consequences of the possible costs and benefits of opening data along each decision branch and identifies decision nodes required in further analysis.

Based on our experimental case studies, we conclude that there is no single best method to support governments, academia, and communities in identifying the potential advantages and disadvantages of opening data. The selection of the methods depends on several factors. The factors might include the divers of the organisational policy and their objectives, lack of human-skilled resources, different levels of formal education, and lack of experience in analysing the datasets.

8.2 Revisiting the research questions

Our study performed a comprehensive systematic literature review approach to answer the first research question (RQ#1. The first research question was defined as *what are the advantages and disadvantages of opening data?* Although there are overviews of advantages and disadvantages in the literature review, there was no structure in taxonomy for supporting decision-making. Therefore, to answer RQ#1, we conducted two main steps of a systematic literature review. First, we explored the definitions and key concepts of open government data. Our study contributes to classifying Open Government Data initiatives into seven perspectives: general overview, political and policy-making, institutional, social and cultures, economics, technical and human's cognitive, and legislation. These classifications aim to become easier to study, understand, compare, and contrast the related properties among the characteristics of open government data domain and their compounds from different groups. Second, we investigated the advantages and disadvantages of opening data by conducting a systematic literature review. Our study found four primary categories to represent the advantages and disadvantages of opening data. The categories include political and legislation, technology, social, and economy. Finally, a taxonomy was developed to help decision-makers define the advantages and disadvantages of opening data.

After we explored the OGD key concepts' definition, we then searched for elements of the decision-making support for opening data to answer the second research question (RQ#2). The second research question was defined as *what are the elements of decision-making support for opening data?* To answer RQ#2, we conducted a comprehensive systematic literature review and a preliminary case study in Indonesia's government institutions. Our comprehensive literature study found that there are eight essential elements to develop our DSOD prototype. The elements include Database Management System (DBMS), Model Base Management system (MBMS), Dialogue Generation and Management System (DGMS), User interface, User authentication, Decision context, Knowledge-based, and The model and analytical tool. Also, from the preliminary case study, the best practice on how decision-makers in Indonesia decide to open data at the time was a valuable input to define the decision-support elements. Our case study showed that three main aspects influenced the decision-making process to open data, namely institutional, technology, and process. By combining the eight elements and the three main influencing aspects in the decision-making process to open data, the answer of RW#2 contributes to the literature regarding the decision-making support requirements and detailed steps that need to be taken in the decision-making process open data.

Our novelty functionalities of the decision-making support to open data were defined to answer the third research question (RQ#3). The third research question was defined as *what are the functionalities of a prototype?* Our study combined the decision-making support elements in the previous part (Chapter 3) and the preliminary case study in Indonesia to define and describe the decision-making support's functionalities to open data. The functionalities of the decision-making support contributed to the development of the DSOD prototype by designing the five main functionalities, namely (1) retrieve and decompose datasets, (2) evaluation the datasets, (3) assessment the datasets, (4) decision-making, and (5) iteration and update the datasets. The functionalities of the DSOD contribute to the literature and practical overview by employing three selected methods (BbN, FMCDM, and DTA).

Furthermore, this study was focused on finding the differences between the three methods used (BbN, FMCDM, and DTA) in the decision-making process for the

OGD stakeholders. It aimed to answer the fifth research question (RQ#4). The fourth research question was defined as *what are the differences between BbN, FMCDM, and DTA to support decision-making about the opening of the dataset?* We performed a quasi-experiment case study using three different participants: government, academia, and community, to answer this research question. The preferences for a method might be dependent on the type of dataset. First, the DTA is the most feasible and preferable method for government institutions in the e-procurement case study. The government actors such as decision-makers, policy-makers, and administrative officers put confidence in employing the DTA since it could be served usefulness for analysing the potential costs and benefits of opening data. Second, the BbN is the best alternative method for government and academia to analyse the causal relationships between factors influencing the advantages and disadvantages of opening datasets both in the e-procurement and medical records case study. Academia gave credence to use BbN as the most feasible method for them because of its ability to show the transparency process in the decision-making process to open data. Third, BbN is also the community's preferred method to analyse the causal relationships between factors influencing the advantages and disadvantages of disclosing data. Community actors such as professional data analytics, non-government organisations, and general users agree to accept the BbN since it provides ease of steps and a better understanding of the decision-making process. The BbN can weigh utilities the disadvantages and advantages integrally against each other and consider uncertainties in cause-effect relationships in the opening e-procurement dataset.

8.3 Why decision-making processes to open data are not trivial?

This dissertation finds many stakeholder tensions in the decision-making process to open data. Our study defines several essential factors of government institutions' challenges as data publishers to decide to open data. First, stakeholders might be different. Some of the stakeholders, such as civil servants, are risk-averse and want to avoid making mistakes, whereas politicians promote transparency and accountability of the government organisations. Different stakeholders can have various concerns, and each stakeholder has their own goals to use and re-analyse datasets. For example, there is a tension between the value of open data at the internal government organisation level and avoiding risks. In contrast, other stakeholders consistently provide new knowledge to the public, while others tend to be risk-averse.

Second, politicians may play a substantial role in determining which objectives dominate and how many resources are allocated for opening datasets. However, there are no clear priorities and goals to guide decision-making. The high political ambitions are in sharp contrast to the limited resources, a lack of adequate infrastructure, a limited number of trained staff, and supporting methods for making decisions.

Third, stakeholders such as decision-makers and data analysts are already overloaded and have limited time resulting in the quick release of the datasets, which often have limited value. The focus is on reaching the target of opening several datasets instead of looking at the societal benefits. For some data, there is a clear 'yes' or 'no' for opening data, but for other datasets, this is less clear and straightforward. The decision-makers and civil servants lack cognitive and limited time to an in-depth analysis of the datasets. It requires training and skills in mathematics and computing, whereas the educational background is often low. Our decision-making support might only be viewed as consuming time and not adopted for this reason.

Fourth, there is a potential of conflicting interests between the data publishers about the extent to which data should be opened and affect decision-making delays. The challenges result in tensions in the level of ambitions, resource allocation, acquiring technical infrastructure, and training. As a consequence, data is all too often not disclosed.

Finally, our study also found several stakeholder tensions in the decision-making process to open datasets. There might be no single answer possible for some datasets as a trade-off between creating societal benefits and risk appetite might be needed. The different roles of stakeholders represent different objectives and concerns. Each stakeholder has its own goals and objectives to use and re-analyse datasets. This situation results in tensions between ambitions and the actual opening of data. In the internal government organisation, there is a tension between the value of opening data and the avoidance of risks. Some stakeholders are consistent with providing new knowledge to the public, while others tend to be risk-averse. There are no clear priorities and objectives to guide decision-making. Stakeholders are overloaded and have limited time, resulting in the quick release of datasets, often having limited value.

8.4 Research limitations

The results of this research have several limitations. First, this study has been carried out from the interpretivist paradigm, which supports multiple insights and realities derived from human beings (Goldkuhl, 2012; Walsham, 1993). The interpretivism approach in this research is essential to understand the decision-making process's circumstances to open data. The cases include the motivation behind the decision-making process to open or close the data. Interpretivism is used to understand the relationships between the advantages and disadvantages of opening data and other subjective experiences that can emerge during the observations and experimental case study. However, interpretative research has been criticized for not thoroughly evaluating objective evaluation criteria and mechanisms (Goldkuhl, 2012; Klein & Myers, 1999). The possible subjectivity from the different stakeholders in determining the advantage and disadvantage factors is a limitation. Also, the subjective assessment by the participants in evaluating the three methods in the quasi-experiment is a limitation.

Second, we focussed on decision-making support, but we did not analyse the political processes resulting in open data policies in the decision-making process. Power, sense of urgency, and dominance (Josefin Lassinantti et al., 2013) can play a role in decision-making.

Third, we focussed on public servants in Indonesia. In other countries, the decision-making process might be different, and persons' attitudes might be different. Therefore the results can be different in other countries.

Finally, the prototype was developed, and the quasi-experiment was conducted with a limited number of participants. More or other participants might yield different outcomes.

8.5 Further research recommendations

In the previous sections of this chapter, we highlighted what has been undertaken in this dissertation, including the study's limitations. This final section provides several research directions for further research agenda in the decision-making support for opening data domain.

Our first recommendation is to investigate other decision-making support methods. There are many methods for deciding to open data, and these methods can yield different objectives and advantages, as our case study experiment showed in Chapter 7. As more and more decisions are made, data providers can collect data about the decision-making process's inputs and outputs. Machine Learning (ML) algorithms and Artificial Intelligence (AI) are the example methods that could be used to process further the opening dataset (Houston, Edge, & Bernier, 2019). In this case, ML and AI can provide an intelligent system to learn from the historical data and its decision experiences automatically. Therefore, in future research, the ideal method needs to be developed with all the best aspects.

The second recommendation is to collect and store information needed to make decisions about the opening datasets already when collecting open datasets. The decision to open datasets depends on the required information details in the context. Such information is required to access the possibility of misuse of the information if privacy is violated. However, we found that these circumstances were not available and that it is time-intensive to collect such datasets. The decision-making process can be efficient and effective if the datasets would preferably be collected already untimely by default and automatically.

Our third recommendation for the research direction is to employ more technical controls to help data publishers make decisions. There might be no single answer possible for some datasets as a trade-off between creating societal benefits and risk appetite that might be required. Simultaneously, several alternative suppression mechanisms like pseudonymization, generalising data fields, and data aggregation could help train the selected datasets.

The prior research studies that are relevant to decision-making support to open data is limited. As a result, our fourth recommendation is to develop a research typology of participant groups and identify gaps in the prior literature. The use of research typology makes the different types of participant groups and their requirements more specific. Regarding the focused methods used in this research, other methods might be relevant or combined to arrive at even better decision support. Hence, we recommend further research in this area.

Finally, our case studies represent the context of the government organisations of Indonesia. We recommend analysing the decision-making process better to understand politics, decision-makers roles, and public servants' interests. The evaluation results of the methods were different in the two quasi-experiments. Therefore, our last recommendation for the research direction is to conduct more experiments in different contexts and settings.

References

Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2019). The Process of Open Data Publication and Reuse. *Journal of the Association for Information Science and Technology, 70*(3), 296-300. doi:DOI: 10.1002/asi.24116

Adina Tofan, C. (2015). Decision Tree Method Applied in Cost-based Decisions in an Enterprise. *Procedia Economics and Finance, 32*, 1088-1092. doi:https://doi.org/10.1016/S2212-5671(15)01572-5

Ahmad Mir, S., & Quadri, S. M. K. (2009). Decision Support Systems: Concepts, Progress and Issues – A Review. *Sustainable Agriculture Reviews, 2*, 373-399.

Aikenhead, G. S. (1997). *A Framework for Reflecting on Assessment and Evaluation Globalization of Science Eduation*. Paper presented at the International Conference on Science Education, Seoul.

Al-Khouri, A. M. (2012). Data Ownerhip: Who owns 'My Data'? *International Journal of Management & Information Technology, 2*(1).

Alamoudi, E., Mehmood, R., Aljudaibi, W., Albeshri, A., & Hamid Hasan, S. (2020). *Open Source and Open Data Licenses in the Smart Infrastructure Era: Review and License Selection Frameworks*.

Albano, C. S., & Reinhard, N. (2014). *Open Government Data: Facilitating and Motivating Factors for Coping with Potential Barriers in the Brazilian Context*. Paper presented at the International Conference on Electronic Government.

Aldossary, S., & Allen, W. (2016). Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions. *International Journal of Advanced Computer Science and Applications, 7*(4), 485-498. doi:10.14569/issn.2156-5570

Alexopoulos , C., Zuiderwijk, A., Charapabidis, Y., Loukis, E., & Janssen, M. (2014). *Designing a Second Generation of Open Data Platforms: Integrating Open Data and Social Media*. Paper presented at the 13th International Conference on Electronic Government (EGOV), Dublin, Ireland.

Ali-Eldin, A. M. T., Zuiderwijk, A., & Janssen, M. (2017). *Opening More Data: A New Privacy Scoring Model of Open Data*. Paper presented at the Seventh International Symposium on Business Modelling and Software Design (BMSD 2017), Barcelona, Spain.

Alonso, M. J. (2011). *Open Government Data (approaches, concerns and barriers, lessons learned)*. Retrieved from Brussel:

Altman, M., Wood, A., O'Brien, D., Vadhan, S., & Gasser, U. (2015). Towards a Modern Approach to Privacy-Aware Government Data Releases. *Berkeley Technological Law Journal, 30*(3). doi: http://dx.doi.org/10.15779/Z38FG17

Alzamil, Z. S., & Vasarhelyi, M. A. (2019). A new model for efective and efcient open government data. *International Journal of Disclosure and Governance*. doi:doi:10.1057/s41310-019-00066-w

Amit, S. P., & Larson, J. A. (1990). Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. *ACM Computing Surveys, 22*(3), 183-236. doi:10.1145/96602.96604

Andanda, P. (2019). Towards a Paradigm Shift in Governing Data Access and Related Intellectual Property Rights in Big Data and Health-Related Research. *International Review of Intellectual Property and Competition Law, 50*, 1052-1081. doi:https://doi.org/10.1007/s40319-019-00873-2

Andreoli-Versbacha, P., & Mueller-Langera, F. (2014). Open access to data: An ideal professed but not practised. *Research Policy, 43*, 1621-1633. doi:http://dx.doi.org/10.1016/j.respol.2014.04.008

Aparicio, M., & Costa, C. J. (2014). Data Visualization. *Communication Design Quarterly, 3*(1), 7-11. doi:10.1145/2721882.2721883

Archer, P., Dekkers, M., Geodertier, S., Hazard, N., & Loutas, N. (2013). *Business models for Linked Open Government Data: what lies beneath?* Retrieved from

Arcidiacono, D., & Reale, G. (2016). Young activists between data disclosure and digital reputation. *The Open Journal of Sociopolical Studies, 9*(3), 919-947. doi:DOI: 10.1285/i20356609v9i3p918

Ariss, A. (2017). Machine-readable Open Data: How it's appicable to developing countries. Retrieved from https://blogs.worldbank.org/opendata/machine-readable-open-data-how-it-s-applicable-developing-countries

Arnott, D., & Pervan, G. (2012). Design Science in Decision Support Systems Research: An Assessment using the Hevner, March, Park, and Ram Guidelines. *Journal of Association for Information Systems, 13*(11), 923-949. doi:10.17705/1jais.00315

Arnott, D., & Pervan, G. (2014). A critical analysis of decision support systems research revisited: the rise of design science. *Journal of Information Technology, 14*, 1-25. doi:10.1057/jit.2014.16

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., . . . Wouters, P. (2004). Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal, 3*.

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A Systematic Review of Open Government Data Initiatives. *Government Information Quarterly, 32*, 399-418. doi:http://dx.doi.org/10.1016/j.giq.2015.07.006

Barnickel, N., Flügge, M., Hanke, D., Höfig, E., Klessmann, J., Marienfeld, F., . . . Ziesing, J. (2012). Berlin Open Data Strategy. In *Concept, pilot system and recommendations for action*: Organisational, legal and technical aspects of Open Data in Berlin.

Barry, E., & Bannister, F. (2014). Barriers to Open Data Release: A View from the Top. *Information Polity, 19*(1-2), 129-152. doi:https://doi.org/10.3233/IP-140327

Bates, J. (2014). The strategic importance of information policy for the contemporary neoliberal state: The case of Open Government Data in the United Kingdom. *Government Information Quarterly, 31*(3), 388-395. doi:https://doi.org/10.1016/j.giq.2014.02.009

Baumeiester, R. F., & Leary, M. R. (1997). Writing Narrative Literature Review. *Review of General Psychology*. doi:https://doi.org/10.1037/1089-2680.1.3.311

Bayazit, N. (2004). Investigating Design: A Review of Forty Years of Design Research. *Design Issues, 20*(1), 16-29. doi:10.1162/074793604772933739

Bayraktarli, Y. Y., & Ulfkjaer, J. P. (2005). *On the Application of Bayesian Probabilistic Networks for Earthquake Risk Management.* Paper presented at the 9th

International Conference on Structural Safety and Reliability (ICOSSAR 05), Rome, Italy 2005.

Beaudouin-Lafon, M., & Mackay, W. (2002). Prototyping tools and techniques. In *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications* (pp. 1006-1031).

Beaudrie, C., Kandlikar, M., & Ramachandran, G. (2016). *Using Expert Judgment for Risk Assessment*

Behkamal, B., Kahani, M., Bagheri, E., & Jeremic, Z. (2013). A Metrics-Driven Approach for Quality Assessment of Linked Open Data. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(2), 64-79. doi:DOI: 10.4067/S0718-18762014000200006

Ben-Gal, I. (2008). *Bayesian Networks: Encyclopedia of Statistics in Quality and Reliability*.

Bernstein, L. (1996). Foreword: Imporatance of Software Prototyping *Journal of System Integration, 6*, 9-14.

Berrang-Ford, L., Pearce, T., & Ford, J. D. (2015). Systematic review approaches for climate change adaptation research. *Regional Environment Change, 15*, 755-769.

Bertot, J. C., Jaeger, P. T., & Grimmes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly, 27*(3), 264-271. doi:doi.org/10.1016/j.giq.2010.03.001

Beuzen, T. (2018). A comparison of methods for discretizing continuous variables in Bayesian Networks. *Environmental Modelling & Software, 108*, 61-66. doi:https://doi.org/10.1016/j.envsoft.2018.07.007

Bichard, J.-A., & Knight, G. (2012). Improving public services through open data: public toilets. *Institution of Civil Engineer*, 157-165.

Biesbroek, R., Berrang-Ford, L., Ford, J. D., Tanabe, A., Austin, S. E., & Lesnikowski, A. (2018). Data, concepts and methods for large-n comparative climate change adaptation policy research: A systematic literature review. *WIREs Climate Change, 9*, 1-15. doi:DOI: 10.1002/wcc.548

Black, I. R. (2006). The Presentation of Intervists Research. *Qualitative Market Research, 9*(4), 319-324. doi:DOI: 10.1108/13522750610689069

Boer, Harink, J., & Heijboer, G. (2002). Aconceptual model for assessing the impact ofelectronic procurement. *European Journal of Purchasing & Supply Management, 8*, 25-33.

Bonczek, R. H., Holsapple, C., & Whinston, A. (1980). The Evolving Roles of Models in Decision Support System. *Decision Science, 11*(2), 337-356. doi:10.1111/j.1540-5915.1980.tb01143.x

Borgesius, Z., Frederik, Z. B., van Eechoud, M., & Gray, J. (2015). *Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework*. Retrieved from

Borghi, J. A., Abrams, S., Lowenberg, D., Simms, S., & Chodacki, J. (2018). Support Your Data: A Research Data Management Guide for Researchers. *Research Ideas and Outcomes*.

Borglund, E., & Engvall, T. (2014). Open Data? Data, information, document or record? *Records Managements Journal, 24*(2), 136-181. doi:DOI 10.1108/RMJ-01-2014-0012

Bøttcher, S. G., & Dethlefsen, C. (2003). *Learning Bayesian Networks with R*, Vienna University of Technology.

Bradley, M. (2014). *Overview of the Intellectual Property Government Open Data*.

Brickell, J., & Shmatikov, V. (2008). *The cost of privacy: Destruction of data-mining utility in anonymized data publishing.* Paper presented at the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Nevada, USA.

Buda, A., Ubacht, J., Warnier, M., Janssen, M., Nagel, R., & Sipis, R. (2015). *Decision support framework for opening business data.* (Master), Delft University of Technology, Delft. Retrieved from https://repository.tudelft.nl/islandora/object/uuid:9033ea68-20ee-4acd-8bd4-91f595d82b5f (uuid:9033ea68-20ee-4acd-8bd4-91f595d82b5f)

Budde, R., Kautz, K., Kuhlenkamp, K., & Züllighoven, H. (1992). What is prototyping? *Information Technology and People, 6*(2/3), 89-95.

Buneman, P., Davidson, S., & Hillebrand, G. (1996). *A query language and optimization techniques for unstructured data*. Paper presented at the ACM SIGMOD International Conference on Management of Data.

Burwell, S. M., VanRoekel, S., Park, T., & Mancini, D. J. (2013). *Open Data Policy - Managing Information As an Asset*. Retrieved from https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

Caimi, C., Gambardella, C., Manea, M., Petrocchi, M., & Stella, D. (2015). *Legal and Technical Perspectives in Data Sharing Agreements Definition.* Paper presented at the Annual Privacy Forum, Luxembourg, Luxembourg.

Campos, A., & Evans, A. M. (2013). Open Government Initiatives: Challanges of Citizen Participation. *Journal of Policy Analysis and Managemen, 32*(1), 172-185.

Canares, M. P., Marcial, D., & Narca, M. (2016). Enhancing Citizen Engagement with Open Government Data. *Journal of Community Informatics, 12*(2).

Cárdenas, I. C., Halman, J. I. M., & Al-Jibouri, S. (2012). *A Bayesian belief networks approach to risk control in construction projects.* Paper presented at the 14th International Conference on Computing and Civil Engineering, Moscow, Russian.

Carr, M., & Verner, J. (1998). Prototyping and Software Development Approaches. *Computer Science*.

Carter, L. (2012). E-Government Utilization: Understanding the Impact of Reputation and Risk. *Electronic Government Research, 8*(1), 83-97. doi:DOI: 10.4018/jegr.2012010105

Catherine, T. A. (2012). Trustworthy Records and Open Data. *The Journal of Community Informatics, 8*(2).

Cavaye, A. L. M. (1996). Case study research: a multi-facated research approach for IS. *Information Systems, 6*, 227-242. doi:https://doi.org/10.1111/j.1365-2575.1996.tb00015.x

Ceballos, B., Lamata, M. T., & Pelta, D. (2017). Fuzzy Multicriteria Decision-MakingMethods: A Comparative Analysis. *International Journal of Intelligent Systems, 32*, 722-738. doi:https://doi:10.1002/int.21873

Chakraborty, S., Mengersen, K., Fidge, C., Ma, L., & Lassen, D. (2016). A Bayesian Network-based customer satisfaction model: a tool for management decisions in railway transport. *Journal of Decision Analytics, 3*(4), 2-24. doi:https://doi.org/10.1186/s40165-016-0021-2

Champbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*: Houghton Mifflin.

Chan, C. M. L. (2013). *From Open Data to Open Innovation Strategies: Creating e-Services Using Open Government Data*. Paper presented at the 46th Hawaii International Conference on System Sciences, Hawaii.

Chandler, J. A. (2007). *Negligence Liability for Breaches of Data Security*.

Charalabidis, Y., Loukis, E., & Alexopoulos , C. (2014). *Evaluating Second Generation Open Government Data Infrastructures Using Value Models*. Paper presented at the 47th Hawaii International Conference on System Science, Hawaii.

Chen, D., & Zhao, H. (2012). *Data Security and Privacy Protection Issues in Cloud Computing*. Paper presented at the International Conference and Privacy Protection Issues in Cloud Computing, Hangzhou, China.

Chen, S.-J., & Hwang, C.-L. (1992). Fuzzy Multiple Attribute Decision Making. In *Lecturer Notes in Economics and Mathematical Systems*.

Cheol Kim, H., & Yong Gim, G. (2015). A Study on Public Data Quality Factors Affecting the Confidence of the Public Data Open Policy. *Journal of Korea Society of IT services, 14*(1), 53-68. doi:DOI: 10.9716/KITS.2015.14.1.053

Chockalingam, S., Pieters, W., Teixeira, A. e., Khakzad, N., & van Gelder, P. (2018). *Combining Bayesian Networks and Fishbone Diagrams to Distinguish Between Intentional Attacks and Accidental Technical Failures*. Paper presented at the The Fifth International Workshop on Graphical Models for Security (GraMSec), Oxford, United Kingdom.

Chockalingam, S., Pieters, W., Teixeira, A. e., & Van Gelder, P. (2017). *Bayesian Network Models in Cyber Security: A Systematic Review*. Paper presented at the 22nd Nordic Conference, NordSec 2017, Tartu, Estonia.

Choe, E., Lee, B., Zhu, H., Henry, N., & Baur, D. (2017). *Understanding self-reflection: how people reflect on personal data through visual data exploration*. Paper presented at the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare, Barcelon, Spain.

Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, S10-S17. doi:10.1016/j.giq.2014.01.003

Costantino, G., Martinelli, F., Matteucci , I., & Petrocchi, M. (2017). *Analysis of Data Sharing Agreement*. Paper presented at the International Conference on Information Systems Security and Privacy, Porto, Portugal.

Crandell, J. L., Voils, C. I., & Sandelwoski, M. J. (2012). Bayesian Approaches to the Synthesis of Qualitative and Quantitative Research Findings. In *Medicine*.

Craveiro, G. d. S., Porto De Albuquerque, J., & Tavares de Santana, M. (2013). *Assessing Open Government Budgetary Data in Brazil*. Paper presented at the The Seventh International Conference on Digital Society.

Creswell, J. W. (2014). *The Selection of a Research Approach. In Research design: Qualitative, quantitative, and mixed methods approaches*: SAGE Publications.

Cucciniello, M., Nasi, G., & Valotti, G. (2012). *Assessing Transparency in Government: Rhetoric, Reality and Desire*. Paper presented at the 45th Hawaii International Conference on System Sciences, Hawaii.

Czarnitzkia, D., Hünermund, P., & Moshgbar, N. (2020). Public Procurement of Innovation: Evidence from a German Legislative Reform. *International Journal of Industrial Organization, 71*. doi:https://doi.org/10.1016/j.ijindorg.2020.102620

Davies, T. (2010). *Open data, democracy and public sector reform: A look at open government data use from data. gov. uk*. Retrieved from

De Los Reyes, A., & Kazdin, A. E. (2008). When the Evidence Says, "Yes, No, and Maybe So": Attending to and Interpreting Inconsistent Findings Among Evidence-Based Interventions. *Current Directions in Psychological Science, 17*(1), 47-51.

Dekkers, M., Loutas, N., Keyzer, M. D., & Goedertier, S. (2014). *Data and Metadata Licensing*. Retrieved from

Delgado-Gómez, D., C.Laria, J., & Ruiz-Hernández, D. (2019). Computerized adaptive test and decision trees: A unifying approach. *Expert Systems with Applications, 117*, 358-366. doi:https://doi.org/10.1016/j.eswa.2018.09.052

Deligiannidis, L., Kochut, K. J., & Sheth , A. P. (2007). *RDF data exploration and visualization*. Paper presented at the ACM Workshop on Cyberinfrastructure: Information Management in eScience, Lisbon, Portugal.

Denney, A. S. (2013). How to Write a Literature Review. *Journal of Criminal Justice Education, 24*(2). doi:DOI: 10.1080/10511253.2012.730617

Driss, O. B., Mellouli, S., & Trabelsi, Z. (2019). From citizens to government policy-makers: Social media data analysis. *Government Information Quarterly, 36*(3), 560-570. doi:https://doi.org/10.1016/j.giq.2019.05.002

Dulong de Rosnay, M., & Janssen, K. (2014). Legal and Institutional Challenges for Opening Data across Public Sectors: Towards Common Policy Solutions. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(3), 1-14. doi:DOI: 10.4067/S0718-18762014000300002

Dunsire, G. (2013). *Granularity in Library Linked Open Data*. Paper presented at the Code4Lib 2013, Chicago, USA.

Duval, E., Hodgins, W., Sutton, S., & Weibel, S. (2002). Metadata Principles and Practicalities. *D-Lib Magazine, 8*(4).

Eckartz M, S., Hofman J, W., & Veenstra, A. F. V. (2014). *A Decision Model for Data Sharing*. Paper presented at the International Conference on Electronic Government, Dublin, Ireland.

Elena, V. L., López , C. D., Paciello , J. M., & Pane, J. I. (2016). *Proposal of a unified BPM model for open data publication*. Paper presented at the Third International Conference on eDemocracy & eGovernment (ICEDEG), Sangolqui, Ecuador.

Elliot, A. C. (2006). *Statistical Analysis Quick Reference Guidebook: With SPSS Examples 1st Edition*: SAGE.

Emrouznejad, A. (2016). *Big Data Optimization: Recent Developments and Challenges*: Springer.

Estermann, B. (2014). Diffusion of Open Data and Crowdsourcing among Heritage Institutions: Results of a Pilot Survey in Switzerland. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(3).

European Commission. (2011). Digital agenda: Commission's open data strategy, questions & answers. Retrieved from http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/891&format=HTML&aged=1&language=EN&guiLanguage=en

European Commission. (2013). Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the Re-use of Public Sector Information. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32013L0037

European Commission. (2019). Directive EU (2019/1024) of the European Parliament and of the Council of 20 June 219 an open data the re-use of public sector information (recast). Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024

Evans, b. J. (2011). Much ado about data ownership *Harvard Journal of Law & Technology, 25*(1).

Fenton, N., & Neil, M. (2012). *Risks Assessment and Decision Analysis with Bayesian Networks*: Noca Raton: CRC Press.

Filip, F. G., Zamfirescu, C.-B., & Ciurea, C. (2017). *Computer-Supported Collaborative Decision-Making* (Vol. 4). Sibiu, Romania: Springer.

Fleisher, C. S. (2008). Using Open Source Data in Developing Competitive and Market Intelligence. *European Journal of Marketing, 42*(7), 852-866. doi:DOI: 10.1108/03090560810877196

Floyd, C. (1984). A Systematic Look at Prototyping. *Approaches to Prototyping*, 1-18. doi:DOI: 10.1007/978-3-642-69796-8_1

Fraenkel, J. R., & Wallen, N. E. (2011). *How To Design And Evaluate Research In Education*: Mcgraw-Hill Education - Europe.

Fran. Ruiz, J., Petrocchi, M., Matteucci, I., Costantino, G., Gambardella, C., & Manea, M. (2016). *A Lifecycle for Data Sharing Agreements: How it Works Out*. Paper presented at the Annual Privacy Forum, Frankfurt, Germany.

Fuller, R. (1999). *Fuzzy Logic and Neural Nets in Intelligent Systems*. Paper presented at the Information System Day, Turku Centre for Computer Science.

Gao, J., Xie, C., & Tao, C. (2016). *Big Data Validation and Quality Assurance – Issuses, Challenges, and Needs*. Paper presented at the IEEE 10th IEEE International Symposium on Service-Oriented System Engineering, Oxford, United Kingdom.

Garbett, A., Linehan, C., Kirman, B., Wardman, J., & Lawson, S. (2011). *Using social media to drive public engagement with open data*. Paper presented at the Digital Engagement, Newcastle, UK.

Geiger, C. P., & Lucke, J. v. (2012). Open Government and (Linked) (Open) (Government) (Data). *JeDEM - eJournal of eDemocracy and Open Government, 4*(3), 265-278.

Giannopoulou, A. (2018). Understanding Open Data Regulation: An Analysis of the Licensing Landscape. In B. v. L. e. al (Ed.), *Open Data Exposed - Information Technology and Law Series* Springer

Gibson, H., Ramwell, S., & Day, T. (2016). Analysis, Interpretation and Validation of Open Source Data. In *Advanced Sciences and Technologies for Security Applications* (pp. 95-110): Springer International Publishing.

Gigler, B.-S., Rahemtulla, H., & Custer, S. (2011). *Realizing the Vision of Open Government Data: Opportunities, Challenges and Pitfalls*. Paper presented at the World Bank- Open Development Technology- ICT for Development, Washington, USA.

Goëta, S., & Davies, T. (2019). The Daily Shaping of State Transparency: Standards, Machine-Readability and the Configuration of Open Government Data Policies. *Science and Technology Studies, 29*(4), 9-30.

Goldkuhl, G. (2012). Pragmatism vs Interpretivism in Qualitative Information Systems Research. *European Journal Information Systems, 21*(2), 135-146. doi:http://dx.doi.org/10.1057/ejis.2011.54

Gonzales-Zapata, F., & Heeks, R. (2015). The Multiple Meanings of Open Government Data: Understanding Different Stakeholders and Their Perspective. *Government Information Quarterly, 32*, 441-452. doi:http://dx.doi.org/10.1016/j.giq.2015.09.001

Gough, D., Oilver, S., & Thomas, J. (2012). *An Introduction to Systematic Reviews*.

Gour, V., Sarangdevot, Singh, T. G., & Sharma, A. (2010). Improve Performance of Extract, Transform and Load (ETL) in Data Warehouse. *International Journal on Computer Science and Engineering, 2*(3), 786-789.

Green, B. (2017). Open Data Privacy. *Berkman Klein Center for Internet & Society, 17*(07).

Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *Journal of Chiropractic Medicine, 5*(3), 101-117.

Greenhalgh, T., & Russel, J. (2009). Evidence-Based Policymaking: A Critique. *Perspectives in Biology and Medicine, 52*(2), 304-318. doi:DOI: 10.1353/pbm.0.0085

Gregory, K. (2020). A dataset describing data discovery and reuse practices in research. *Scientific Data, 7*(232). doi:https://doi.org/10.6084/m9.figshare.12445034

Grimmelikhujsen, S. G., & Meijer, A. J. (2014). Effects of Transparency on the Perceived Trustworthiness of a Government Organization: Evidence from an Online Experiment. *Journal of Public Administration Research and Theory, 24*(1), 137-157.

Guba, E. G., & Lincoln, Y. (1982). Epistemological and methodological bases of naturalistic inquiry. *Educational Technology Research and Development, 30*(8), 233-252.

Gupta, M. M. (1995). Fuzzy Logic and Neural Systems. *International Series in Intelligent Technologies*, 225-244.

Gurin, J. (2014). Open Governments, Open Data: A New Lever for Transparency, Citizen Engagement, and Economic Growth. *SAIS Review of International Affairs, 34*(1), 71-82. doi:DOI: 10.1353/sais.2014.0009

Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday, 16*(2). doi:10.5210/fm.v16i2.3316 · Source: DBLP

Haag, S., Cummings, M., & Dawkins, J. (2000). *Management Information Systems for the Information Age*: Irwin/McGraw-Hill.

Hancerliogullari, G., Oymen, K. H., & Koksalmis, E. (2017). The use of multi-criteria decision making models in evaluating anesthesia method options in circumcision surgery. *BMC Medical Informatics and Decision Making, 17*(14). doi:10.1186/s12911-017-0409-5

Harzing, A.-W., & Alangkas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics, 106*(2), 787-804. doi:https://doi.org/10.1007/s11192-015-1798-9

Head, B. (2010). *Evidence-based policy: principles and requirements*, University of Queensland.

Heckerman, D. (2008). A Tutorial on Learning With Bayesian Networks. In D. Holmes (Ed.), *Innovation in Bayesian Networks* (pp. 33-82): Springer-Verlag Berlin.

Helbig, N., Cresswell, A. M., Burke, B. G., & Luna-Reyes, L. F. (2012). *The Dynamics of Opening Government Data*. Retrieved from Albany, New York:

Herland, K., Hämmäinen, H., & Kekolahti, P. (2016). Information Security Risks Assessment of Smartphones Using Bayesian Networks. *Journal of Cyber Security, 4*, 65-85. doi:https://doi.org/10.13052/jcsm2245-1439.424

Hevner, A. R., & Chatterjee, S. (2010). Design research in Information Systems. *New York: Springer*.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information System Research. *MIS Quarterly, 28*(1), 75-105.

Hielkema, H., & Hongisto, P. (2013). Developing the Helsinki Smart City: The Role of Competitions for Open Data Applications. *Journal of Knowledge Economic, 4*, 190-204. doi:DOI 10.1007/s13132-012-0087-6

Hoffmann, S. (2016). Bridging the Gap between Perception and Cognition: An Overview. *Performance Psychology*, 135-149.

Holden, M. T., & Lynch, P. (2004). Choosing the Appropriate Methodology: Understanding Research Philosophy. *The Marketing Review, 4*(4), 2-18. doi:DOI: 10.1362/1469347042772428

Honda, H., Washida, Y., Sudo, A., Wajima, Y., & Awata, K. (2017). The Difference in Foresight Using the Scanning Method Between Experts and Non-Experts. *Technological Forecasting and Social Change, 119*, 18-26. doi:https://doi.org/10.1016/j.techfore.2017.03.005

Horn, W., Miksch, S., Egghart, G., Popow, C., & Paky, F. (1998). Effective data validation of high-frequency data: Time-point-, time-interval-, and trend-based methods. *Computers in Biology and Medicine, 27*(5), 389-409. doi:https://doi.org/10.1016/S0010-4825(97)00012-7

Horný, M. (2014). *Bayesian Networks*. Retrieved from Boston University School of Public Health:

Houston, P., Edge, V., & Bernier, E. (2019). Reaping the benefits of Open Data in public health. *Canada Communicable Disease Report, 45*(11), 252-256. doi:10.14745/ccdr.v45i10a01

Hsieh, T.-Y., Lu, S.-T., & Tzeng, G.-H. (2004). Fuzzy MCDM Approach for Planning and Design Tenders Selection in Public Office Buildings. *International Journal of Project Management, 22*, 573-584. doi:doi:10.1016/j.ijproman.2004.01.002

Hudson, L. A., & Ozanne, J. L. (1988). Alternatives Ways of Seeking Knowledge In Consumer Research. *Journal of Consumer Research, 14*(4), 508-521.

Hudson, L. D., Ware, B. S., Mahoney, S. M., & Laskey, K. B. (2002). An Application of Bayesian Networks to Antiterrorism Risk Management for Military Planners. *Online*.

Husmeir, D. (2005). *Introduction to Learning Bayesian Networks from Data*. London: Springer.

Idreos, S., Papaemmanouil, O., & Chaudhuri, S. (2015). *Overview of Data Exploration Techniques*. Paper presented at the ACM SIGMOD International Conference on Management Data.

Isaak, A. C. (1981). Generalisation in Political Science. *Scope and Methods of Political Science*, 105-132.

Isselhardt, D. J., & Cappuci, J. (1989). The Analytic Hierarchy Process in Medical Decision Making A Tutorial. *Med Decis Mak, 1*, 40-50.

Ivanov, M., Varga, M., & Bach, M. P. (2014). *Government open data portal: How government strategies should be more open*. Paper presented at the 7th International Conference: An Enterprise Odyssey: Leadership, Innovation and Development for Responsible Economy, , Zadar , Croatia.

Jan, B., Farman, H., Khan, M., Imran, M., & Islam, I. (2019). Deep Learning in Big Data Analytics: A comparative Study. *Computers and Electronic Engineering, 75*, 275-287. doi:https://doi.org/10.1016/j.compeleceng.2017.12.009

Janssen, K. (2011). The influence of the PSI directive on open government data: An overview ofrecent developments. *Government Information Quarterly, 28*, 446-456. doi:http://dx.doi.org/10.1016/j.giq.2011.01.004

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information System Management, 29*(4), 258-268. doi:10.1080/10580530.2012.716740

Janssen, M., & Van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in Government: A Challange to transparency and privacy? . *Government Information Quarterly, 32*(4), 363-368.

Jetzek, T. (2016). Managing Complexity across Multiple Dimensions of Liquid Open Data. *Government Information Quarterly, 33*(1), 89-104. doi:DOI: 10.1016/j.giq.2015.11.003

Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2014). Data-Driven Innovation through Open Government Data. *journal of Theory Application and Electronic Commerce, 9*(2). doi:http://dx.doi.org/10.4067/S0718-18762014000200008

Jetzek, T., Bjorn-Andersen, N., & Avital, M. (2013). *Generating Value from Open Government Data*. Paper presented at the Thirty Fourth International Conference on Information Systems, Milan, Italy.

Johnston, K. A. (2007). *Community Engagement: A Relational Perspective*. Paper presented at the Communications, Civics, Industry – ANZCA2007.

Juana-Espinosa, S. (2020). Open government data portals in the European Union: A dataset from 2015 to 2017. *Data in Brief, 29*. doi:https://doi.org/10.1016/j.dib.2020.105156

Julia Zhu, Y., & Freund, L. *Exploring Open Data Initiatives in Higher Education*. Paper presented at the 15th International Conference, iConference 2020, Boras, Sweden.

Kahraman, C., Onar, S. C., & Oztaysi, B. (2015). Fuzzy Multicriteria Decision-Making: A Literature Review. *International Journal of Computational Intelligence System, 8*(4), 637-666.

Kaplan, B., & Maxwell, J. A. (2005). *Qualitative Research Methods for Evaluating Computer Information Systems*.

Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicagoopendata project. *Government Information Quarterly, 30*, 508-513. doi:https://doi.org/10.1016/j.giq.2013.05.012

Keim, D. A. (2001). Visual exploration of large data sets. *Communication of the ACM, 44*(8), 39-44. doi:https://doi.org/10.1145/381641.381656

Khusro, S., Jabeen, F., Rahman Mashwani, S., & Alam, I. (2014). Linked Open Data: Towards the Realization of Semantic Web- A Review. *Indian Journal of Science and Technology, 7*(6), 745-764. doi:DOI: 10.17485/ijst/2014/v7i6.7

Kim, Chung, J., & Trimi, S. (2014). Big Data Applications in the Government Sector: A Comparative Analysis among Leading Countries. *Communication of the ACM, 57*(3), 78-85. doi:DOI: 10.1145/2500873

Kim, S., & Chung, Y. D. (2019). An anonymization protocol for continuous and dynamic privacy-preserving data collection. *Future Generation Computer Systems, 93*, 1065-1073. doi:https://doi.org/10.1016/j.future.2017.09.009

King, W., & He, J. (2006). Understanding the role and methods of meta-analysis in IS research. *Communications of the Association for Information Systems, 16*(16).

Kivunja, C., & Kuyini, A. B. (2017). Understandning and Applying Research Paradigms in Educational Contexts. *International Journal of Higher Education, 6*(5), 26-41. doi:doi:10.5430/ijhe.v6n5p26

Klabi, H., Mellouli, S., & Rekik, M. (2018). A reputation based electronic government procurement model. *Government Information Quarterly, 35*, 43-53. doi:http://dx.doi.org/10.1016/j.giq.2016.01.001

Klein, H. K., & Myers, M. D. (1999). A Set of Principles for Conducting and Evaluating Interpretative Field Studies in Information Systems. *MIS Quarterly, 23*(1), 67-94.

Knol, A., Slottje, P., Sluijs, J. P. v. d., & Lebret, E. (2010). The use of expert elicitation in environmental health impact assessment: a seven step procedure. *Environmental Health, 9*(19), 1-16. doi:https://doi:10.1186/1476-069X-9-19

Komarova, T., Nekipelov, D., & Yakovlev, E. (2018). Identification, data combination, and the risk of disclosure. *Quantitative Economics, 9*, 395-440. doi:https://doi.org/10.3982/QE568

Kontkanen, P., Myllymäki, P., Silander, T., & Tirri, H. (1997). *Comparing Predictive Inference Methods for Discrete Domains*. Paper presented at the Sixth

International Workshop on Artificial Intelligence and Statistics, Ft. Lauderdale, USA.
Koot, M. R., Noordende, G., & Laat, C. D. (2010). *A study on the re-identifiability of Dutch citizens.* Paper presented at the HotPETs 2010.
Korba, L., Wang, Y., Geng, L., Song, R., Yee, G., Patrick, A. S., . . . You, Y. (2008). *Private Data Discovery for Privacy Compliance in Collaborative Environments.* Paper presented at the International Conference on Cooperative Design, Visualization and Engineering.
Kostkova, P., Brewer, H., Lusignan, S., Fottrell, E., Goldacre, B., Hart, G., . . . Tooke, J. (2016). Who Owns the Data? Open Data for Healthcare. *Frontiers of Public Health, 4*(7), 1-6. doi:10.3389/fpubh.2016.00007
Kothari, C. (2004). *Research Methodology: Methos & Techniques*: New Age International Limited.
Koutsoukis, N.-S., Dominguez-Ballesteros, B., Lucas, C., & Mitra, G. (2000). A prototype decision support system for strategic planning under uncertainty. *International Journal of Physical Distribution & Logistics Management, 30*(7/8), 640-661. doi:DOI: 10.1108/09600030010346387
Krotoski, A. K. (2012). Data-driven research: open data opportunities for growing knowledge, and ethical issues that arise. *Sociology, Computer Science*. doi:DOI:10.1629/2048-7754.25.1.28
Kucera, J., & Chlapek, D. (2014). Benefits and Risks of Open Data. *Journal of System Integration, 1*, 30-41. doi:10.20470/jsi.v5i1.185
Kučera, J., Chlapek, D., & Nečaský, M. (2013). *Open Government Data Catalogs: Current Approaches and Quality Perspective.* Paper presented at the International Conference on Electronic Government and the Information Systems Perspective, Prague, Czech Republic.
Kuk, G., & Davies, T. (2011). *The roles of agency and artifacts in assembling open data complementarities.* Paper presented at the 32nd International Conference on Information Systems (ICIS), Shanghai, China.
Kulk, S., & van Loenen, B. (2012). Brave New Open Data World? *International Journal of Spatial Data Infrasructure Research, 7*, 196-206. doi:10.2902/1725-0463.2012.07.art10
Kuo, M., S, Liang, G. S., & Huang, W. C. (2006). Extension of the multicriteria analysis with pairwise comparison under a fuzzy environment. *Journal of Science Direct, 43*, 268-285.
Kupzyk, K. A., & Cohen, M. Z. (2015). Data Validation and Other Strategies for Data Entry. *Western Journal of Nursing Research, 37*(4), 546-556. doi:DOI: 10.1177/0193945914532550
Kutner, M., & Wasserman, W. (1996). *Applied Linear Statistical Models*: Mcgraw-Hill Education - Europe.
Lamy, J.-B., Ellini, A., Nobécourt, J., Venot, A., & Zucker, J.-D. (2010). *Testing Methods for Decision Support Systems*: Open Access.
Lassinantti, J., Bergvall-Kåreborn, B., & Ståhlbröst, A. (2013). Shaping Local Open Data Initiatives: Politics and Implications. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(2), 17-33. doi:10.4067/S0718-18762014000200003

Lassinantti, J., Bergvall-Kåreborn, B., & Ståhlbröst, A. (2014). Shaping local open data initiatives: politics and implications. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(9). doi:https://doi.org/10.4067/S0718-18762014000200003

Lee, G., & Kwak, Y. H. (2012). An Open Government Maturity Model for social media-based public engagement. *Government Information Quarterly, 29*. doi:http://dx.doi.org/10.1016/j.giq.2012.06.001

Lee, M.-H. (2019). Challenges of Designing with Open Data: The Case of Cultural Data. *Archives of Design Research, 32*(3), 57-72. doi:DOI: 10.15187/adr.2019.08.32.3.57

Liang, J. (2012). *Government Cloud: Enhancing Efficiency of E-Government and Providing Better Public Services*. Paper presented at the 2012 International Joint Conference on Service Sciences, Shanghai, China.

Licklider, J. C. R. (1960). Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics, 1*, 4-11.

Lin, C., & Twu, C. H. (2012). Fuzzy MCDM for Evaluating Fashion Trend Alternatives. *International Journal of Clothing Science and Technology, 24*(2/3), 141-153. doi:https://doi.org/10.1108/09556221211205586

Linders, D. (2012). From E-Government to We-Government: Defining a Typology for Citizen Coproduction in the Age of Social Media. *Government Information Quarterly, 29*(4), 446-454. doi:10.1016/j.giq.2012.06.003

Linders, D. (2013). Towards open development: Leveraging open data to improvethe planning and coordination of international aid. *Government Information Quarterly, 30*, 426-434. doi:http://dx.doi.org/10.1016/j.giq.2013.04.001

Lindman, J. (2014). Similarities of Open Data and Open Source: Impacts on Business. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(3). doi:http://dx.doi.org/10.4067/S0718-18762014000300006

Little, C. M., & Cooke, R. M. (2016). Expert Judgment and Uncertainty Quantification for Climate Change. *Nature Climate Change, 6*, 445-451.

Liu, K., Chen, C.-W., Lu, C.-F., & Shen, Y.-S. (2012). Applying Bayesian Belief Networks to Health Risks Assessment. *Stochastic Environmental Research and Risk Assessment, 26*(3), 451-465. doi:10.1007/s00477-011-0470-z

Lourenço, R. P. (2013). *Open Government Portals Assessment: A Transparency for Accountability Perspective*. Paper presented at the EGOV 2013, Koblenz, Germany.

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly, 32*(3), 323-332. doi:https://doi.org/10.1016/j.giq.2015.05.006

Lundqvist, B. (2016). *Big Data, Open Data, Privacy Regulations, Intellectual Property and Competition Law in an Internet of Things World*: Faculty of Law, University of Stockholm.

Luthfi, A., & Janssen, M. (2017). *A Conceptual Model of Decision-making Support for Opening Data*. Paper presented at the 7th International Conference, E-Democracy 2017, Athens, Greece.

Luthfi, A., & Janssen, M. (2019a). *A Comparative Study of Methods for Deciding to Open Data*. Paper presented at the Business Modelling and Software Design (BMSD) 2019, Lisbon.

Luthfi, A., & Janssen, M. (2019b). Open Data for Evidence-based Decision-making: Data-driven Government Resulting in Uncertainty and Polarization. *International Journal on Advanced Science Engineering and Information Technology, 9*(3), 1071-1078. doi:DOI: 10.18517/ijaseit.9.3.8846

Luthfi, A., Janssen, M., & Crompvoets, J. (2018a, 2-4 July 2018). *A Causal Explanatory Model of Bayesian-belief Networks for Analysing the Risks of Opening Data.* Paper presented at the 8th International Symposium, BMSD 2018, Vienna, Austria.

Luthfi, A., Janssen, M., & Crompvoets, J. (2018b). *Framework for Analyzing How Governments Open Their Data: Institution, Technology, and Process Aspects Influencing Decision-Making.* Paper presented at the EGOV-CeDEM-ePart 2018, Donau-Universität Krems, Austria.

Luthfi, A., Janssen, M., & Crompvoets, J. (2019). *Decision Tree Analysis for Estimating the Costs and Benefits of Disclosing Data.* Paper presented at the 18th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2019, Trondheim, Norway.

Luthfi, A., Janssen, M., & Crompvoets, J. (2020). *Stakeholder Tensions in Decision-making for Opening Government Data.* Paper presented at the Business Modelling and Software Design (BMSD) 2020, Berlin, Germany.

Luthfi, A., Rehena, Z., Janssen, M., & Crompvoets, J. (2018). *A Fuzzy Multi-criteria Decision Making Approach for Analyzing the Risks and Benefits of Opening Data.* Paper presented at the 17th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2018, Gulf University for Science and Technology (GUST), Kuwait.

Luthfi, A., Rukanova, B., Molenhuis, M., Janssen, M., & Tan, Y.-H. (2020). *Bayesian-belief Networks for Supporting Decision-making of the Opening Data by the Customs* Paper presented at the EGOV-CeDEM-ePart 2020, Linköping University, Sweden.

Mach, K., Mastrandrea, M., Freeman, P., & Field, C. (2017). Unleashing expert judgment in assessment. *Global Environmental Change, 44*, 1-14. doi:https://doi.org/10.1016/j.gloenvcha.2017.02.005

Máchová, R. (2007). *Measuring the Effects of Open Data on the Level of Corruption.* Paper presented at the 21th International Conference Current Trends in Public Sector Research.

Magalhaes, G., Roseira , C., & Strover, S. (2013). *Open Government Data Intermediaries: A Terminology Framework* Paper presented at the 7th International Conference on Theory and Practice of Electronic Governance, Seoul, South of Korea.

Malamud, C., Elin, G., Sifry, M., Holovaty, A., O'Neil, D. X., Migurski, M., . . . Lessig, L. (2007). Open Government Data Principles. Retrieved from https://public.resource.org/8_principles.html

Mallach, E. (1994). *Understanding Decision Support Systems and Expert Systems*: McGraw-Hill School Education Group.

Manheim, J. B., & Rich, R. C. (1995). Theory Building: Concepts and Hypotheses in Political Science. *Empirical Political Analysis*.

Manyika, J., Chui, M., Groves, P., Farrel, D., Kuiken, S. V., & Doshi, E. A. (2013). *Open data: Unlocking innovation and performance with liquid information*. Retrieved from

Marakas, G. H. (1999). *Decision Support Systems in the Twenty-first Century*: Prentice Hall.

March, S. T., & Smith, G. (1995). Design and natural science research on information technologies. *Decision Support Systems, 15*(4), 251-266.

Marsh, W. E. (1993). Decision Tree Analysis: Drawing Some of the Uncertainty Out of Decision Making. *Swine Health Production, 1*(4), 17-23.

Marshall, A., Brynjolfsson, E., & Madnick, S. (1995). Why not one big database? Principles for data ownership. *Decision Support Systems, 15*.

Marshall, C., & Rossman, G. B. (2010). *Designing Qualitative Research*: SAGE Publisher.

Martin, S., Foulonneau, M., Turki, S., & Ihadjadene, M. (2013). Risk Analysis to Overcome Barriers to Open Data. *Electronic Journal of e-Government, 11*(1), 348-359.

Matthews, B., & Wensveen, S. (2015). *Prototypes and prototyping in design research* (J. Y. Paul A. Rodgers Ed.): Routledge

Mayernik, M. S. (2017). Open data: Accountability and transparency. *Big Data & Society*. doi:10.1177/2053951717718853

Mayinka, J., Chui, M., Groves, P., Farrel, D., Kuiken, S. V., & Doshi, E. A. (2013). *Open data: Unlocking innovation and performance with liquid information*. Retrieved from

McDermott, P. (2010). Building Open Government. *Government Information Quarterly, 27*, 401-413. doi:10.1016/j.giq.2010.07.002

McDonald, J., & Léveillé, V. (2014). Whither the retention schedule in the era of big data and open data? *Records Managements Journal, 24*(2), 99-121. doi:DOI 10.1108/RMJ-01-2014-0010

Meijer, A., & Thaens, M. (2009). Public information strategies: Making government information available to citizens. *Information Polity, 14*(1-2), 31-45. doi:10.3233/IP-2009-0167

Mitra-Kahn , B., Johnson , M., Man, B., & Meehan, L. (2016). Intellectual Property Government Open Data: Australian Business Number Links to All Intellectual Property Data in Australia. *The Australian Economic Review, 49*(1), 96-104.

Mittal, K., & Khanduja, D. (2017). An Insight into "Decision Tree Analysis". *World Wide Journal of Multidisciplinary Research and Development, 3*(12), 111-115.

Mockus, M., & Palmirani, M. (2015). *Open Government Data Licensing Framework*. Paper presented at the Electronic Government and the Information Systems Perspective, Valencia, Spain.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine*. doi:https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Mohsen, D., Shafri, M., Zulhaidi, H., Noordin, A., Pradhan, B., & Sahabeh, S. (2014). A Combined Fuzzy MCDM Approach for Identifiying the Suitable Lands for

Urban Development: An Example from Bandar ABBS, Iran. *Journal of Urban and Environmental Engineering, 8*(1), 11-27. doi:http://10.4090/juee.2014.v8n1.011027

Monroe, P. A. (2011). Evidence-Based Policymaking: Insights from Policy-Minded Researchers and Research-Minded Policymakers. *journal of Marriage and Family, 73*(3), 687-689. doi:DOI: 10.2307/29789612

Murphy, K. (1998). *A Brief Introduction to Graphical Models and Bayesian Networks*. Retrieved from

Myers, M. D. (1997). Qualitative Research in Information Systems. *MIS Quarterly, 21*(2). doi:DOI: 10.2307/249422

Nadkarni, S., & Shenoy, P. P. (2001). A Bayesian network approach to making influence in causal maps. *European Journal of Operational Research, 128*, 479-498.

Napolitano, M. (2019). *Open data as a public asset*. Retrieved from

Nath, R., & Pavur, R. (1985). A new statistic in the one way multivariate analysis of variance. *Computational Statistics and Data Analysis, 2*, 297-315.

Neopolitan, R. E. (2003). *Learning Bayesian Networks*. Northeastern Illinois University: Prentice Hall Series in Artificial Intelligence.

Neuman, L. (2014). *Social Research Methods: Qualitative and Quantitative Approaches*: Pearson Education Limited.

Nielsen, J. (2012). Usability 101: Introduction to Usability. *Human Computer Interaction, User Testing, Web Usability*.

Nkurunziza, S. e. e. (2019). *Efficiently Combining Data from Various Sources*. Paper presented at the The 13th International Conference on Management Science and Engineering Management, St. Catharines, Ontario, Canada.

Noble, E. (2020). Data Types. In (pp. 3-23): Apress, Berkeley, CA.

Nojava, F., Qian, S., & Stow, C. (2017). Comparative analysis of discretization methods in Bayesian networks. *Environmental Modelling & Software, 87*, 64-71. doi:https://doi.org/10.1016/j.envsoft.2016.10.007

Norrington, L., Quigley, J., Russell, A., & Van der Meer, R. (2008). Modelling the Reliability of Search and Resque Operations with Bayesian Belief Networks. *Reliability Engineeding & System Safety, 93*(7), 940-949. doi:https://doi.org/10.1016/j.ress.2007.03.006

Novák, K. (1992). An Introduction to Fuzzy Logic Applications in Intelligent System. *Kluwer Academic*, 185-200.

Nunamaker, J. F., Chen, M., & Purdin, T. D. M. (1991). System Development in Information System Research. *Journal of Information System, 7*(3), 89-106.

Okediji, R. L. (2014). Government as Owner of Intellectual Property? Considerations for Public Welfare in the Era of Big Data. *Vanderbilt Journal of Entertainment & Technology Law, 18*(8), 331.

Omidvar, M., & Bordbar, F. (2013). Advanced decision support systems for managers. *European Online Journal of Natural and Social Sciences, 2*(3), 700-708.

Open Knowledge Foundation. (2015). Open Definition 2.1. Retrieved from https://opendefinition.org/od/2.1/en/

Oslon, K. (2010). An Examination of Questionnaire Evaluation by Expert Reviewers. *Field Methods, 22*(4), 295-318.

Ozair, F. F., Jamshed, N., Sharma, A., & Aggarwal, P. (2015). Ethical issues in electronic health records: A general overview. *Perspective in Clinical Research, 6*(2), 73-76. doi:10.4103/2229-3485.153997

Paré, G., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management, 52*(3), 183-199.

Parycek, P., Hochtl, J., & Ginner, M. (2014). Open Government Data Implementation Evaluation. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(2). doi:http://dx.doi.org/10.4067/S0718-18762014000200007

Pasquetto, I. V., Randles, B. M., & Borgman, C. (2017). On the Reuse of Scientific Data. *Data Science Journal, 16*(8), 1-9. doi:http://doi.org/10.5334/dsj-2017-008

Pearl, J., & Russel, S. (2001). *Bayesian Networks*. Retrieved from Cambridge:

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems, 24*(3), 45-78.

Peignot, J., Peneranda, A., Amabile, S., & Marcel, G. (2013). Strategic Decision Support Systems for Local Government: A Performance Management Issue?

The Use of Information Systems on the Decision-making and Performance Management of Local Government. *International Business Research, 6*(2), 92-100. doi:10.5539/ibr.v6n2p92

Pereira, G. V., Macadar, M. A., Luciano, E. M., & Testa, M. G. (2017). Delivering public value through open government data initiatives in a Smart City context. *Information System Frontier, 19*, 213-229. doi:DOI 10.1007/s10796-016-9673-7

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology, 64*, 1-18. doi:https://doi.org/10.1016/j.infsof.2015.03.007

Piotrowski, S. J., & Ryzin, G. G. V. (2007). Citizen Attitudes Toward Transparency in Local Government. *The American Review of Public Administration, 37*(3), 306-323.

Pliskin, N., & Shoval, P. (1987). End-user Prototyping: Sophisticated Users Supporting System Development. *Database Summer, 7*.

Podvezko, V. (2011). Application of AHP Technique. *Journal of Business Economics and Management, 10*(2), 181-189. doi: https://doi.org/10.3846/1611-1699.2009.10.181-189

Pomerantz, J. (2015). *Metadata*. Cambridge: MIT Press.

Power, J. D. (2002). *Decision Support Systems: Concepts and Resources for Managers*: University of Northern Iowa.

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2014). *Artifact Evaluation In Information Systems Design-Science Research - A Holistic View*. Paper presented at the PACIS 2014 Proceedings.

Puron-Cid, G., Gil-Garcia, R. J., & Luna-Reyes, L. F. (2012). *IT-enabled policy analysis: new technologies, sophisticated analysis and open data for better government decisions*. Paper presented at the 13th Annual International Conference on Digital Government Research, College Park, Maryland, USA. Proceedings retrieved from

Putri Nugroho, R., Zuiderwijk, A., Janssen, M., & de Jong, M. (2015). A comparison of national open data policies: Lessons learned. *Transforming Government: People, Process and Policy, 8*(17). doi:http://dx.doi.org/10.1108/TG-03-2014-0008

Quinlan. (1990). Decision Trees and Decision-making. *IEEE Transactions on Systems, Man, and Cybernetics, 20*(2), 339-346. doi:10.1109/21.52545

Raffaghelli , J. E., & Manca, S. (2019). Is There a Social Life in Open Data? The Case of Open Data Practices in Educational Technology Research. *Publications - MDPI Open Access Journal*. doi:doi:10.3390/publications7010009

Rajshree, N., & Srivastava, B. (2012). Open Government Data for Tackling Corruption – A Perspective. *Semantic Cities*.

Reeves, T. C., & Hedberg, J. G. (2003). *Interactive Learning System Evaluation*: Educational Technology Publications.

Reggi, L. (2011). Benchmarking Open Data Availability across Europe: The Case of EU Structural Funds. *European Journal of ePractice*.

Rezaei, P., Rezaei, K., Nazari-Shirkouhi, S., & Tajabadi, M. R. J. (2013). Application of Fuzzy Multi-Criteria Decision Making Analysis for Evaluating and Selecting the Best Location for Construction of Underground Dam. *Acta Polytechnica Hungarica, 10*(7), 187-205.

Rhoades, E. A. (2011). Literature Reviews. *The Volta Review, 111*(1), 61-71.

Robertson, D., & Wang, Q. J. (2004). Bayesian networks for decision analyses - An application to irrigation system selection. *Australian Journal of Experimental Agriculture, 44*(2). doi:10.1071/EA02178

Rocher, L., M. Hendrickx, J., & Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *nature Communication*. doi:https://doi.org/10.1038/s41467-019-10933-3

Rowlands, B. (2005). Grounded in Practice: Using Interprete Research to Build Theory. *The Electronic Journal of Business Research Methdology, 3*(1), 81-92. doi:DOI: 10.1.1.134.9432

Roy, C., Swarup Rautaray, S., & Pandey, M. (2018). Big Data Optimization Techniques: A Survey. *International Journal of Information Engineering and Electronic Business, 4*(41), 41-48. doi:DOI: 10.5815/ijieeb.2018.04.06

Rush, C., & Roy, R. (2001). Expert judgement in cost estimating: Modelling the reasoning p. *Concurent Engineering, 9*, 271-284. doi:https://doi.org/10.1177/1063293X0100900404

Saaty, T. (1980). The Analytic Hierarchy Process Planning, Priority Setting, Resource Allocation. *New York: McGraw*.

Safarov, I., Grimmelikhuijsen, S., & Meijer, A. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity, 22*(1), 1-24. doi:DOI: 10.3233/IP-160012

Salini, S., & Kennet, R. (2007). Bayesian networks of customer satisfaction survey data. *Journal of Applied Statistics, 36*(2007-33). doi:1080/02664760802587982

Sapsford, R., & Jupp, V. (2006). *Data Collection and Analysis*.

Saunders, M., Lewis, P., & Thornhill, A. (2007). *Research Methods for Business Students*: Pearson Education Limited.

Sauter, V. L. (2010). *Decision Support Systems for Business Intelligence*: John Wiley & Sons, Inc. All rights reserved.

Saxena, S., & Muhammad, I. (2017). The impact of open government data on accountability and transparency. *Journal of Economic and Administration Sciences*. doi:DOI 10.1108/JEAS-05-2017-0044

Scasa, T. (2014). Privacy and Open Government. *Jornal of Future Internet, 6*(2), 397-413. doi:10.3390/fi6020397

Scholl, H. J., & Luna-Reyes, L. F. (2011). *Transparency and openness in government: A system dynamics perspective.* Paper presented at the 5th International Conference on Theory and Practice of Electronic Governance, Talinn, Estonia.

Schwegmann, C. (2012). *Open Data in Developing Countries*. Retrieved from

Sehra, S. K., Brar, Y. S., & Kaur, N. (2012). Multi Criteria Decision Making Approach for Selecting Effort Estimation Model. *International Journal of Computer Application, 39*(1), 10-17.

Serra, C. (2014). The mapping, selecting and opening of data. *Records Managements Journal, 24*(2), 87-98. doi:DOI 10.1108/RMJ-01-2014-0008

Simon, H. A. (1972). Theories of Bounded Rationality. *Decision and Organization*.

Simon, H. A. (1979). Rational Decision Making in Business Organizations. *The American Economic Review, 69*(4), 493-513.

Sloane, E., Liberatore, M., & Nydick, R. (2011). Medical Decision Support Using the Analytic Hierarchy Process. *Journal of Health Informatics Management, 16*(4), 38-43.

Smith, M. F. (1991). *Software Prototyping: Adoption, Practice, and Management*: McGraw-Hill.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333-339.

Sommerville, I. (2011). *Software Engineering*: Addison-Wesley.

Song, Y.-y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry, 27*(2), 130-135. doi:https://doi.org/10.11919/j.issn.1002-0829.215044

Spiegelhalter, D. J. (1998). Bayesian graphical modelling: a case-study in monitoring health outcomes. *Royal Statistical Society, 47*(1), 115-133. doi:10.1111/1467-9876.00101

Spooner, L. M., & Pesaturo, K. A. (2013). Chapter 1: The Medical Record. In.

Sprague, R. H., & Carlson, E. D. (1982). *Building Effective Decision Support Systems*: Prentice Hall Professional Technical Reference.

Stone, H., Bleibaum, R., & Thomas, H. (2020). Chapter 6 - Descriptive analysis. In *Sensory Evaluation Practices (Fifth Edition)*.

Straßheim, H., & Kettunen, P. (2014). When does evidence-based policy turn into policy-based evidence? Configurations, contexts, and mechanisms. *Evidence & Policy A Journal of Research Debate and Practice, 10*(2), 259-277. doi:DOI: 10.1332/174426514X13990433991320

Strydom, W., Nortje, K., Funke, N., & Steyn, M. (2010). Evidence-based policymaking: A review. *South African Journal of Science, 106*(5/6), 1-8. doi:DOI: 10.4102/sajs.v106i5/6.249

Stutz, J., & Cheesman, P. (1994). A Short Exposition on Bayesian Inference and Probability. *National Aeronautic and Space Administration Ames Research Centre: Computational Sciences Division, Data Learning Group*.

Sun, Y., Yongping, X., Zhang, J., & Zhu, G. (2014). Data Security and Privacy in Cloud Computing. *International Journal of Distributed Sensor Networks, 6*(50), 1-9. doi:10.1155/2014/190903

Sundell, K., Tengvald, K., Soydan, H., & Anttila, S. (2009). From Opinion-Based to Evidence-Based Social Work: The Swedish Case. *Research on Social Work Practice, 19*(5). doi:http://10.1177/104973150934788

Sunderberg, L. (2016). Decision Making and Value Realization in Multi-Actor e-Government Context. *Electronic Government and Electronic Participation, 23*, 147-154. doi:10.3233/978-1-61499-670-5-147

Susha, I., Grönlund, Å., & Janssen, M. (2015). Driving factors of service innovation using open government data: An exploratory study of entrepreneurs in two countries. *Information Polity, 20*(1), 19-34.

Sutcliffe, S., & Court, J. (2005) Evidence-Based Policymaking: What is it? How does it work? What relevance for developing countries? In.

Swarup, V., Seligman, L., & Rosenthal, A. (2006). *A Data Sharing Agreement Framework*. Paper presented at the Information Systems Security, Second International Conference, Kolkata, India.

Sweeney, L. (2002). Achieving k-anonimity Privacy Protection Using Generalization and Suppression. *International Journal on Ucertainty, Fuzziness and Knowledge-based System, 10*(5), 571-588.

Sylvester, A., Tate, M., & Johnstone, D. (2010). Beyond synthesis: re-presenting heterogeneous research literature. *Behaviour & Information Technology, 32*(12), 1199-1215. doi:https://doi.org/10.1080/0144929X.2011.624633

Tabachnick, B., & Fidell, L. (2011). Multivariate Analysis of Variance (MANOVA). *International Encyclopedia of Statistical Science*, 902-904. doi:doi:10.1007/978-3-642-04898-2_394

Talha, M., Abou El Kalam, A., & Elmarzouqi, N. (2019). *Big Data: Trade-off between Data Quality and Data Security.* Paper presented at the The 9th International Symposium on Frontiers in Ambient and Mobile Systems (FAMS) Leuven, Belgium.

Tanaka, H. (2016). Open Data as an Infrastructure: Impact of Availability of Government Data as Open Data on the Japanese Economy. *Policy Research Institute, 12*(1), 23-45.

Taneja, H., Kapil, & Singh, A. K. (2015). *Preserving Privacy of Patients based on Re-identification Risk* Paper presented at the International Conference on Eco-friendly Computing and Communication Systems, Haryana, India.

Teicher, M. (2015). *Interviewing Subject Matter Experts*. Paper presented at the International Cost Estimating and Analysis Association (ICEAA).

Thyer, B. A. (2012). *Quasi-Experimental Research Designs*: Oxford University Press.

Todorov, V., & Filzmoser, P. (2010). Robust statistic for the one-way MANOVA. *Computational Statistics and Data Analysis, 54*(1), 37-448. doi:https://doi.org/10.1016/j.csda.2009.08.015

Toots, M., McBride, K., Kalvet, T., & Krimmer, R. (2017). *Open Data as Enabler of Public Service Co-creation: Exploring the Drivers and Barriers*. Paper presented at the Conference for E-Democracy and Open Government (CeDEM), Krems, Austria.

Tran, E., & Scholtes, G. (2015). *Open Data Literature Review*. Retrieved from University of California:

Truli, E. (2018). *The General Data Protection Regulation and Civil Liability*: Springer-Verlag GmbH Germany.

Tsiavos, P., Karounos, T. G., & Stefaneas, P. S. (2013). The Transposition of European Union Open Data/Public Sector Information Policies in Greece: A Critical Analysis. *Policy and Internet, 5*(4). doi:DOI: 10.1002/1944-2866.POI347

Turban, E., Aronson, J. E., & Liang, T.-P. (2007). *Decision Support System and Intelligent System*: Prentice Hall.

Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. *OECD Working Papers on Public Governance, 22*, 60. doi:10.1787/19934351

Uhlir, P. F. (2009). *The Socioeconomic Effects of Public Sector Information on Digital Networks: Toward a Better Understanding of Different Access and Reuse Policies*. Paper presented at the National Research Council, Washington DC.

Urahn, S. K., Caudell-Feagan, M., & Stasch, J. (2014). *Evidence-Based Policymaking: A guide for effective government*. Retrieved from

van Aken, J. E., & Romme, G. (2012). A Design Science Approach to Evidence-Based Management. In D. M. Rousseau (Ed.), *Design science: research methodology in entrepreneurship and innovation studies* (pp. 43-57).

van Loenen, B., Vancauwenberghe, G., Crompvoets, J., & Dalla Corte, L. (2018). *Open Data Exposed*.

Vassiliadis, P., & Simitsis, A. (2009). Extraction, Transformation, and Loading. *Encyclopedia of Database Systems*, 1-7. doi:https://doi.org/10.1007/978-0-387-39940-9_158

Veen, D., Stoel, D., Zondervan-Zwijnenburg, & Schoot, R. v. d. (2017). Proposal for a Five-Step Method to Elicit Expert Judgment. *Frontiers in Psychology, 8*, 1-11. doi:htttps://doi: 10.3389/fpsyg.2017.02110

Veenstra, A. F. V., & Broek, T. A. v. d. (2013). *Opening Moves – Drivers, Enablers and Barriers of Open Data in a Semi-public Organization*. Paper presented at the International Conference on Electronic Government, Porto, Portugal. Lecturer Notes in Computer Science retrieved from

Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly, 31*, 278-290. doi:http://dx.doi.org/10.1016/j.giq.2013.10.011

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and applicationto Open Government Data. *Government Information Quarterly, 33*, 625-337. doi:http://dx.doi.org/10.1016/j.giq.2016.02.001

Vir Singh, P., & Phelps, C. (2007). Determinants of Open Source Software License Choice: A Social Influence Perspective *Journal Contribution*.

Volmink, J. (2017). Evidence-informed Policy Making: Challanges and Opportunities. *British Medical Journal Global Health, 2*(2). doi:DOI: 10.1136/bmjgh-2016-000260.3

Walker, K. D., Catalano, P., Hammitt, J. K., & Evans, J. S. (2003). Use of expert judgment in exposure assessment: Part 2. Calibration of expert judgments about personal exposures to benzene. *Journal Of Exposure Analysis And Environmental Epidemiology, 13*, 1. doi:10.1038/sj.jea.7500253

Walsham, G. (1993). *Interpreting Information Systems in Organizations*.

Walter, S. (2001). Heterogeneous Database Integration in Biomedicine. *Journal of Biomedical Informatics, 34*(4), 285-298. doi:10.1006/jbin.2001.102

Wang, H.-J., & Lo, J. (2016). Adoption of open government data among government agencies. *Government Information Quarterly, 33*, 80-88. doi:http://dx.doi.org/10.1016/j.giq.2015.11.004

Wang, V., & Shepherd, D. (2020). Exploring the extent of openness of open government data – A critique of open government datasets in the UK. *Government Information Quarterly, 97*(1), 1-10. doi:https://doi.org/10.1016/j.giq.2019.101405

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly, 26*(2), 13-23.

Weikum, G. (2013). Data Discovery. *Data Science Journal, 12*(12), 26-31. doi:DOI: 10.2481/dsj.GRDI-005

Werro, N. (2015). Fuzzy Classification of Online Customers. *Fuzzy Management Methods*. doi:10.1007/978-3-319-15970-6_2

Williams, G. (2018). *Applied Qualitative Research Design*: ED-TECH PRESS.

Wilson, A. G., Graves, T. L., Hamada, M. S., & Reese, C. S. (2006). Advances in Data Combination, Analysis and Collection for System Reliability Assessment. *Statistical Science, 21*(4), 514-531. doi:DOLI: 10.1214/088342306000000439

Witarsyah Jacob, D., Farhan, M., Fudzee, M., Aizi Salamat, M., Kasim, S., Mahdin, H., & Azhar Ramli, A. (2017). Modelling End-User of Electronic-Government Service: The Role of Information quality, System Quality and Trust. *Materials Science and Engineering, 226*. doi:doi:10.1088/1757-899X/226/1/012096

Wu, M., Psomopoulos, F., Jodha Khalsa, S., & de Waard, A. (2019). Data Discovery Paradigms: User Requirements and Recommendations for Data Repositories. *Data Science Journal, 18*(3), 1-13. doi:DOI: 10.5334/dsj-2019-003

Xyntarakis, M., & Antoniou, C. (2019). Data Science and Data Visualization. In *Mobility Patterns, Big Data and Transport Analytics* (pp. 107-144).

Yang, Z., & Kankanhalli, A. (2013). *Innovation in Government Services: The Case of Open Data*. Paper presented at the Innovation in Government Services: The Case of Open Data, Bangalore, India.

Yannoukakou, A., & Araka, I. (2014). Access to Government Information: Right to Information and Open Government Data Synergy. *3rd International Conference on Integrated Information (IC-ININFO), 147*, 332-340. doi:10.1016/j.sbspro.2014.07.107

Yeoa, B., & Grant, D. (2018). Predicting service industry performance using decision tree analysis *International Journal of Information Management, 38*(1), 288-300. doi:https://doi.org/10.1016/j.ijinfomgt.2017.10.002

Yoose, B., & Perkins, J. (2019). The Linked Open Data Landscape in Libraries and Beyond. *Journal of Library Metadata, 13*(2-3). doi:DOI: 10.1080/19386389.2013.826075

Yuanyuan, P., Derek, B. A., & Bob, L. (2018). Rockburst prediction in kimberlite using decision tree with incomplete data. *Journal of Sustainable Mining, 17*, 158-165. doi:https://doi.org/10.1016/j.jsm.2018.07.004

Zadeh, L. A. (1975). The Concept of Lingustic Variable and Its Application to Approximate Reasoning. *Information Sciences, 8*(3), 199-249. doi:https://doi.org/10.1016/0020-0255(75)90036-5

Zeleti , F. A., Ojo, A., & Curry, E. (2016). Exploring the economic value of open government data. *Government Information Quarterly, 33*, 535-551. doi:http://dx.doi.org/10.1016/j.giq.2016.01.008

Zhang, J. (2012). Two-way MANOVA with Unequal Cell Sizes and Unequal Cell Covariance Matrices. *Technometrics*. doi:DOI: 10.1198/TECH.2011.10128

Zhang, S. (2012). Decision Tree Classifier Sensitive to Heterogeneous Costs. *Journal of Systems and Software, 85*(4), 771-779. doi:https://doi.org/10.1016/j.jss.2011.10.007

Zhou, G., & Wang, L. (2012). Co-location decision tree for enhancing decision-making of pavement maintenance and rehabilitation. *Transportation Research Part C, 21*, 287-305.

Zuiderwijk, A. (2017). Analysing Open Data in Virtual Research Environments: New Collaboration Opportunities to Improve Policy Making. *Information Polity, 13*(4), 76-92. doi:DOI: 10.4018/IJEGR.2017100105

Zuiderwijk, A., Choenni, S., Janssen, M., & Meijer, R. (2014). Design Principles for Improving the Process of Publishing Open data. *Transforming Government People Process and Policy, 8*(2). doi:DOI: 10.1108/TG-07-2013-0024

Zuiderwijk, A., Gascó, M., Parycek, P., & Janssen, M. (2014). Special issue on transparency and open data policies: guest editors' introduction. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(3). doi:http://dx.doi.org/10.4067/S0718-18762014000300001

Zuiderwijk, A., & Janssen, M. (2013a). *A Coordination Theory Perspective to Improve the Use of Open Data in Policy-Making*. Paper presented at the IFIP EGOV 2013, Berlin Heidelberg.

Zuiderwijk, A., & Janssen, M. (2013b). Open Data Policies, Their Implementation and Impact: A Framework for Comparison. *Government Information Quarterly, 31*(1). doi:10.1016/j.giq.2013.04.003

Zuiderwijk, A., & Janssen, M. (2014). *Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View*: Springer.

Zuiderwijk, A., & Janssen, M. (2015). Towards decision support for disclosing data: Closed or open data? *Information Polity, 20*(2-3), 103-107. doi:https://doi.org/10.3233/IP-150358

Zuiderwijk, A., & Janssen, M. (2017). *The negative effects of open government data - investigating the dark side of open data*. Paper presented at the the 15th Annual International Conference on Digital Government Research, Mexico City.

Zuiderwijk, A., Janssen, M., & David, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity, 19*(2-3), 17-33. doi:10.3233/IP-140329

Zuiderwijk, A., Janssen, M., Meijer, R., Choenni, S., Charalabidis, Y., & Jeffery, K. (2012). *Issues and Guiding Principles for Opening Governmental Judicial Research Data*. Paper presented at the 11th International Conference on Electronic Government (EGOV), Kristiansand, Norway.

Zuiderwijk, A., Jeffery, K., & Janssen, M. (2012a). *The necessity of metadata for open linked data and its contribution to policy analyses.* Paper presented at the E-Democracy and Open Government (CeDEM12, Krems, Austria.

Zuiderwijk, A., Jeffery, K., & Janssen, M. (2012b). The potential of metadata for linked open data and its value for users and publishers. *eJournal of eDemocracy and Open Government (JeDEM), 4*(2), 222-244.

Zuiderwijk, A., Jeffery, K., & Janssen, M. (2013). *Towards an e-infrastructure to support the provision and use of open data.* Paper presented at the e-Democracy and Open Government (CeDEM13), rems an der Donau, Austria.

Appendices

Appendix A. Preliminary Case Study

Interview Form			
Province	:		
Interviewee	:		
Interviewer	:		
Location	:		
Date & Time	:		
Survey section used	:	☐ A. Interview background	☐ B. Institutional perspective
		☐ C. Design and Strategy	☐ D. Practical Measurement

A. Interview Background

A1	How long have you been working in this institution?	:	
A2	How long have you been in the present position?	:	
A3	What is your highest degree?	:	
A4	What is your field of study?	:	
A5	Briefly describe your role as it relates to your present position?	:	

B. Institutional Perspective

B1	What is the strategy of this institution for improving datasets services for society? Is it working? Why or why not? Probe: To what extent is the enthusiasm of the society in accessing the resources of the datasets?	:	
B2	What resources are available to this institution for improving datasets services for society? Probe: To what extent is the readiness of human resources, infrastructure and applications developed by the Government?	:	

B3	Have you or your colleagues encountered resistance to open datasets reforms in your department or institution? Probe: To what extent is internal organisational support in for opening data?	:
B4	How do you decide to open data? Probe: To what extent is the method used to open the data effective?	:
B5	What benefits does this institution acquire at the planning and experience stage when opening datasets to the public? Probe: To what extent is the method used to open the data effective?	:

C. Design and Strategy

C1	Do this institution has implemented the concept of open government data? Since when? Are there any referral documents from an authorized official? *Probe:* To what extent is the role of the central Government in supporting this program?	:
C2	What types of datasets would like to open to the public by this institution? Which is the full access? Partly access? or Closed? *Probe:* What are the procedures and methods of collecting raw data to the relevant departments? How to select the full, partly, and closed the datasets?	:
C3	What types of datasets have not been opened yet to the public or society by this institution? *Probe:* To what extent is responsibility for datasets that have not been or have been released?	:
C4	What factors may cause the reluctant decision to open the datasets to the public? Why? *Probe:* Is there a difference of interest among officials of decision-makers in this institution?	:

C5	If your department is required to accelerate the implementation of open data in the near future, what is the strategy to be taken? *Probe:* What about the readiness of human resources, finance, and IT infrastructure that exist today?	:
C6	How are you involved in Decision-making on the publishing of government data? *Probe:* Have you ever had a dilemma to weigh the eligibility for data publication?	:

D. Practical Measurements

D1	How do you and your related divisions been collecting, selecting, categorising, and weighing the datasets until deciding which datasets can be opened to the public?	:
D2	When the datasets have been decided upon for publication by your institution, is there a post-issue surveillance procedure for monitoring the datasets?	:
D3	Does your institution already use the supporting system to weigh the potential risks and benefits of open data implementation? Probe: How does the supporting system work?	:
D4	Did any incidents relate to the risks of open data that potentially affects to your institution? Probe: How do tracking mechanisms against data appear to have potential risks when opened to the public?	:
D5	How is the evaluation mechanism regarding open datasets to the public or society? Probe: How effective is the evaluation system implemented so far?	:
D6	How is the evaluation mechanism for opening datasets to the public or society? Probe: How effective is the evaluation system implemented so far?	:

E. Check list of Potential Risks and Benefits

1	Anonymity	☐	1	Availability and access	☐	1	Accessibility	☐
2	Confidentiality	☐	2	Bureaucracy	☐	2	Accountability	☐
3	Errors identification	☐	3	Closed government culture	☐	3	Better social services	☐
4	Heterogonous	☐	4	Conflict of interest	☐	4	Community engagement	☐
5	Inaccuracy	☐	5	Cost impacts	☐	5	Credibility	☐
6	Incomplete metadata	☐	6	Data comparability	☐	6	Data exchange	☐
7	Integration	☐	7	Data compatibility	☐	7	Data-driven decisions	☐
8	Integration	☐	8	Data divers format	☐	8	Innovation	☐
9	License	☐	9	Data linking and combining	☐	9	New application	☐
10	Manipulation	☐	10	Data provider interaction	☐	10	New business opportunity	☐
11	Misuse	☐	11	Data quality	☐	11	Preserves information	☐
12	Ownership	☐	12	Data scalability	☐	12	Public education	☐
13	Privacy	☐	13	Data transformation	☐	13	Public trust	☐
14	Security	☐	14	Digital divide	☐	14	Reputation	☐
15	Sensitivity	☐	15	Disparity of datasets	☐	15	Return of investment	☐
16	Synchronizatio n	☐	16	Free access vs convenient	☐	16	Skilled in employment	☐
17	User skills	☐	17	Human resource skills	☐	17	Transparency	☐
18	Violation	☐	18	Infrastructure development	☐			
19	Vulnerability	☐	19	Jurisdictional	☐			
			20	Lack of business model	☐			
			21	Multiple languages required	☐			

	22	No unified data structures	☐
	23	Officials reluctant	☐
	24	Storing big data	☐
	25	Unclear data provenance	☐
	26	Understandabi lity	☐

Appendix B. Experimental Case Study

I **Demography**
Dear participant, The objective of this pre-test survey is to collect data about your personal information and experiences in using Open Government Data (OGD). This survey consists of 2 main parts and 11 questions in total. Completing the survey will take about 6-8 minutes. You will be asked to fill out another survey after performing this preliminary survey.

1	Please indicate your gender	☐ Male ☐ Female
2	Please indicate your age	☐ 18 – 24 ☐ 25 – 34 ☐ 35 – 44 ☐45 – 54 ☐ >54
3	What is the highest degree or level of education you have completed?	☐ Bachelor degree ☐ Master's degree ☐ Doctorate degree
4	What best describes your employment status?	☐ Executive board
		☐ Policy-maker
		☐ Government employee
		☐ Politician
		☐ Data analyst
		☐ Software system developer
		☐ PhD candidate
		☐ Bachelor or Master student
		☐ Independent researcher
		☐ Full or associate professor
		☐ Lecturer or senior researcher
		☐ Post-doctoral researcher
		☐ Other, please specify:
5	What type of organisation or agency do you work for?	☐ Local government institution
		☐ Ministry or governmental department
		☐ Community-based organisation
		☐ Non-profit organisation
		☐ University or academy
		☐ Business or private organisation
		☐ Other, please specify:
6		☐ Coordinate or administer one or more programs

	What is your current job function? (please mark all that apply)	☐ Supervise those who provide services to the public
		☐ Deliver specialized or technical services
		☐ Provide services directly to the public
		☐ Deliver specialized or technical services
		☐ Provide administrative support
		☐ Responsible for deciding the public policies
		☐ Analyse and investigate dataset from OGD portal
		☐ Other, please specify:
7	How many years have you been in the current position?	☐ 1 - 5 ☐ 6 - 10 ☐11 - 15 ☐ >15
Experience in using open government dataset		
8	In daily work, how often have you been involved in using or analysing open government data?	☐ Daily or multiple times per day
		☐ Weekly or a few times per week
		☐ Monthly or a few times per month
		☐ Yearly or a few times per year
9	How many years have you been involved in using or analysing open government data?	☐ 1 - 5 ☐ 6 - 10 ☐11 - 15 ☐ >15 ☐ none
10	To what extent do you have experience with open government data	☐ Fundamental awareness (basic knowledge)
		☐ Novice (limited experience)
		☐ Intermediate (practical application)
		☐ Advanced (applied theory)
		☐ Expert (recognised authority)
11	What are your reasons for using open government data? (please mark all that apply)	☐ For my individual research
		☐ To perform a statistical analysis
		☐ To perform policymaking research
		☐ To perform investigations
		☐ For political decision-making
		☐ For combining datasets
		☐ For daily operation in work

I sincerely thank you and appreciate your time in participating in this preliminary survey. Subsequently, please go to the next part of this survey.

II **Performing Scenario**
Dear participant, The next part of this session concern performing scenarios. The objective of this survey is to collect information about your experiences in using a prototype of the decision-support system (DSOD) to open data. There are three main scenario tasks, namely Bayesian-belief Networks, Fuzzy Multi-criteria Decision-making, and Decision Tree Analysis. The scenarios consist of 41 questions in total, and it will take about 90 minutes. Many thanks in advance for conducting these scenario tasks.

Initialisation Process

	To which extent do you find it easy or difficult to complete the following scenario tasks? Please first complete the tasks shown in the column on the left. Then answer the question about how easy/difficult this was. If a task is not possible in your opinion then select 'very difficult'.					
		Very difficult	Difficult	Neither easy nor	Easy	Very easy
1	Task 1. *User authentication*. Enter username and password and click "Sign in" button to start the DSS prototype.	☐	☐	☐	☐	☐
2	Task 2. *Message inbox*. Read any incoming messages from the system administrator and click "Continue" button.	☐	☐	☐	☐	☐
3	Task 3. *Dataset selection*. Select a dataset category using the provided combo menu. Health patient records	☐	☐	☐	☐	☐
4	Task 4. *Dataset format*. Select the expected dataset format using radio button (CSV, XML, JSON, HDF)	☐	☐	☐	☐	☐
5	Task 5. *Dataset source*. Find the dataset source or directory at the local computer by clicking "Browse" button. *c:/localcomputer/desktop/dataset/health_patient_records/*	☐	☐	☐	☐	☐
6	Task 6. *Extract dataset*. Click the "Extract dataset" button to decompose the structure of the selected	☐	☐	☐	☐	☐

	dataset. Find the metadata and attribute information.					
7	Task 7. Dataset Information. Find the metadata information of selected dataset.					
8	Task 8. Dataset Information. Find the attributes information of selected dataset.					
9	To which extent were you able to complete tasks 1 to 8? (mark all that apply) ☐ I was able to complete task 1 ☐ I was able to complete task 2 ☐ I was able to complete task 3 ☐ I was able to complete task 4 ☐ I was able to complete task 5 ☐ I was able to complete task 6 ☐ I was able to complete task 7 ☐ I was able to complete task 8 ☐ I was not able to complete any of these tasks					

Scenario 1 **Bayesian-belief Networks**

	Step 1. Analysing Dataset To which extent do you find it easy or difficult to complete the following scenario tasks to analyse dataset? Please first complete the tasks shown in the column on the left. Then answer the question about how easy or difficult this was. If a task is not possible in your opinion then select 'very difficult'.					
		Very difficult	Difficult	Neither easy nor difficult	Easy	Very easy
1	Task 1. *Determine the potential costs and benefits factors*. Select the multiple factors of the costs and benefits for the selected dataset.	☐	☐	☐	☐	☐
2	Task 2. *Construct causal networks*. Click the "Continue" button to generate causal networks of the costs factor.	☐	☐	☐	☐	☐
3	Task 3. *Construct causal networks*. Click the "Continue" button to generate causal networks of the benefits factor.	☐	☐	☐	☐	☐

4	Task 4. *Formalize the structure*. Click the "Continue" button to process the constructed formulation of the Bayesian-belief networks.	☐	☐	☐	☐	☐
5	Task 5. *Quantify the structure*. Fill the three probabilities of the potential costs factor ("High", "Moderate", and "" Low") and click the "Continue" button.	☐	☐	☐	☐	☐
6	Task 6. *Quantify the structure*. Fill the three probabilities of the potential benefits factor ("High", "Moderate", and "" Low") and click the "Continue" button.	☐	☐	☐	☐	☐
7	Task 7. *Interrogate the structure*. Click the "Continue" button to construct the structure of the costs factor.	☐	☐	☐	☐	☐
8	Task 8. *Interrogate the structure*. Click the "Continue" button to construct the structure of the benefits factor.	☐	☐	☐	☐	☐
9	To which extent were you able to complete tasks 1 to 8? (mark all that apply) ☐ I was able to complete task 1 ☐ I was able to complete task 2 ☐ I was able to complete task 3 ☐ I was able to complete task 4 ☐ I was able to complete task 5 ☐ I was able to complete task 6 ☐ I was able to complete task 7 ☐ I was able to complete task 8 ☐ I was not able to complete any of these tasks					
	Step 2. Weighing process To which extent do you find it easy or difficult to complete the following tasks to weigh the potential costs and benefits? Please first complete the tasks shown in the column on the left. Then answer the question about how easy or difficult this was. If a task is not possible in your opinion then select 'very difficult'.					
		Very difficult	Difficult	Neither easy nor difficult	Easy	Very easy

1	Task 1. *Assessment costs and benefits*. Click the "Continue" button to process the costs and benefits assessment works.	☐	☐	☐	☐	☐
2	Task 2. *Matrix comparison*. Click the "Continue" button to process the comparison matrix of costs and benefits factor	☐	☐	☐	☐	☐
3	Task 3. *Pairwise comparison*. Click the "Continue" button to process the result of pairwise comparison.	☐	☐	☐	☐	☐
4	To which extent were you able to complete tasks 1 to 3? (mark all that apply) ☐ I was able to complete task 1 ☐ I was able to complete task 2 ☐ I was able to complete task 3 ☐ I was not able to complete any of these tasks					
	Step 3. Decision made To which extent do you find it easy or difficult to complete the following scenario tasks related to making a decision? Please first complete the tasks shown in the column on the left. Then answer the question about how easy or difficult this was. If a task is not possible in your opinion then select 'very difficult'.					
		Very difficult	Difficult	Neither easy nor difficult	Easy	Very easy
1	Task 1. *Alternative decisions*. Click the "Continue" button to get information about to decision alternatives.	☐	☐	☐	☐	☐
2	Task 2. *Final result*. Click the "Continue" button to get the final result of the Bayesian-belief Networks process.	☐	☐	☐	☐	☐
3	Task 3. Click the "Return" button to loop the analysing process from the beginning.	☐	☐	☐	☐	☐
4	To which extent were you able to complete tasks 1 to 3? (mark all that apply) ☐ I was able to complete task 1 ☐ I was able to complete task 2 ☐ I was able to complete task 3 ☐ I was not able to complete any of these tasks					

Scenario 2 **Fuzzy Multi-criteria Decision Making**	

	Step 1. Structure Decision Alternatives To which extent do you find it easy or difficult to complete the following scenario tasks? Please first complete the tasks shown in the column on the left. Then answer the question about how easy or difficult this was. If a task is not possible in your opinion then select 'very difficult'.					
		Very difficult	Difficult	Neither easy nor difficult	Easy	Very easy
1	Task 1. *Determine the criteria and sub-criteria*. Select the criteria and sub-criteria of potential costs and benefits factor, and click "Continue" button.	☐	☐	☐	☐	☐
2	Task 2. *Construct criteria relationship*. Click the "Continue" button to construct criteria and sub-criteria relationship.	☐	☐	☐	☐	☐
3	Task 3. *Construct criteria hierarchy*. Click the "Continue" button to construct criteria and sub-criteria hierarchy.	☐	☐	☐	☐	☐
4	Task 4. *Define the decision alternative*. Select decision alternatives for each sub-criteria and click "Continue" button to construct decision structure.	☐	☐	☐	☐	☐
5	To which extent were you able to complete tasks 1 to 4? (mark all that apply) ☐ I was able to complete task 1 ☐ I was able to complete task 2 ☐ I was able to complete task 3 ☐ I was able to complete task 4 ☐ I was not able to complete any of these tasks					
	Step 2. Structure Pairwise Comparison To which extent do you find it easy or difficult to complete the following scenario tasks related to weighing the potential costs and benefits factor? Please first complete the tasks shown in the column on the left. Then answer the question about how easy or difficult this was. If a task is not possible in your opinion then select 'very difficult'.					

		Very difficult	Difficult	Neither easy nor difficult	Easy	Very easy
1	Task 1. *Define the fuzzy linguistic scale*. Select the linguistic scales for each criterion and click "Continue" button.	☐	☐	☐	☐	☐
2	Task 2. *Pairwise the criteria*. Select the linguistic scale for each criterion to compare and click "Continue" button.	☐	☐	☐	☐	☐
3	Task 3. *Pairwise the sub-criteria*. Select the linguistic scale for each sub-criterion to compare and click "Continue" button.	☐	☐	☐	☐	☐
4	Task 4. *Calculate the fuzzy weights matrix*. Quantify the fuzzy weights matrix to calculate the decision rank and click "Continue" button	☐	☐	☐	☐	☐
3	To which extent were you able to complete tasks 1 to 4? (mark all that apply) ☐ I was able to complete task 1 ☐ I was able to complete task 2 ☐ I was able to complete task 3 ☐ I was able to complete task 4 ☐ I was not able to complete any of these tasks					

	<u>Step 3. Decision made</u> To which extent do you find it easy or difficult to complete the following scenario tasks related to making the decision? Please first complete the tasks shown in the column on the left. Then answer the question about how easy or difficult this was if a task is not possible in your opinion then select 'very difficult'.

		Very difficult	Difficult	Neither easy nor difficult	Easy	Very easy
1	Task 1. *Alternative decisions*. Click the "Continue" button to get information related to alternative decisions of the selected dataset.	☐	☐	☐	☐	☐
2	Task 2. *Rank decision*. Select the most feasible decision based on the constructed rank of the alternative	☐	☐	☐	☐	☐

	decisions and click the "Continue" button.					
3	Task 3. *Final result*. Click the "Continue" button to get the final result of the fuzzy multi-criteria decision-making process.	☐	☐	☐	☐	☐
4	Task 4. Click the "Return" button to loop the analysing process from the beginning.	☐	☐	☐	☐	☐
5	To which extent were you able to complete tasks 1 to 3? (mark all that apply) ☐ I was able to complete task 1 ☐ I was able to complete task 2 ☐ I was able to complete task 3 ☐ I was able to complete task 4 ☐ I was not able to complete any of these tasks					

Scenario 3 **Decision Tree Analysis**

	<u>*Step 1. Structure Decision Model*</u> To which extent do you find it easy or difficult to complete the following scenario tasks? Please first complete the tasks shown in the column on the left. Then answer the question about how easy or difficult this was. If a task is not possible in your opinion then select 'very difficult'.					
		Very difficult	Difficult	Neither easy nor difficult	Easy	Very easy
1	Task 1. *Determine the costs and benefits factor*. Select the potential costs and benefits factor, and click "Continue" button.	☐	☐	☐	☐	☐
2	Task 2. *Construct a decision structure*. Click the "Continue" button to construct decision structure.	☐	☐	☐	☐	☐
3	To which extent were you able to complete tasks 1 to 2? (mark all that apply) ☐ I was able to complete task 1					

	☐ I was able to complete task 2 ☐ I was not able to complete any of these tasks					
	Step 2. Quantify the Payoffs Variable To which extent do you find it easy or difficult to complete the following scenario tasks related to weighing the potential costs and benefits factor? Please first complete the tasks shown in the column on the left. Then answer the question about how easy or difficult this was. If a task is not possible in your opinion then select 'very difficult'.					
		Very difficult	Difficult	Neither easy nor difficult	Easy	Very easy
1	Task 1. *Assign the payoffs table*. Quantify the probability and investment of the costs and benefits factor and click "Continue" button.	☐	☐	☐	☐	☐
2	Task 2. *Construct a possible consequence.* Click the "Continue" button to construct the possible consequence of the costs and benefits factor.	☐	☐	☐	☐	☐
3	To which extent were you able to complete tasks 1 to 2? (mark all that apply) ☐ I was able to complete task 1 ☐ I was able to complete task 2 ☐ I was not able to complete any of these tasks					
	Step 3. Decision made To which extent do you find it easy or difficult to complete the following scenario tasks related to making the decision? Please first complete the tasks shown in the column on the left. Then answer the question about how easy or difficult this was If a task is not possible in your opinion then select 'very difficult'.					
		Very difficult	Difficult	Neither easy nor difficult	Easy	Very easy
1	Task 1. *Possible paths and total payoffs*. Click the "Continue" button to get information related to possible paths and total payoffs.	☐	☐	☐	☐	☐

3	Task 2. *Final result*. Click the "Continue" button to get the Expected Monetary Value (EMV) and the final result of the Decision Tree Analysis process.	☐	☐	☐	☐	☐
3	Task 3. Click the "Return" button to loop the analysing process from the beginning.	☐	☐	☐	☐	☐
4	To which extent were you able to complete tasks 1 to 3? (mark all that apply) ☐ I was able to complete task 1 ☐ I was able to complete task 2 ☐ I was able to complete task 3 ☐ I was not able to complete any of these tasks					

III **Acceptance of DSOD**
Dear participant, You have just completed the performing scenarios. The next part of this session concerns the acceptance of the proposed DSS model and comparison between the selected methods. This survey consists of 2 main parts and 24 questions in total. Completing the survey will take about 30 minutes. Many thanks in advance for conducting these scenario tasks.

	Please rank (1 to 3) the following in order of the best. *1 = First-best method* *2 = Second-best method* *3 = Third-best method*	
1	Process transparency	(___) Bayesian-belief Networks (___) Fuzzy Multi-criteria Decision-making (___) Decision Tree Analysis
2	Accuracy	(___) Bayesian-belief Networks (___) Fuzzy Multi-criteria Decision-making (___) Decision Tree Analysis
3	Easiness	(___) Bayesian-belief Networks (___) Fuzzy Multi-criteria Decision-making (___) Decision Tree Analysis
4	Usefulness	(___) Bayesian-belief Networks (___) Fuzzy Multi-criteria Decision-making

		(___) Decision Tree Analysis
5	Clear instructions (step-by-step)	(___) Bayesian-belief Networks (___) Fuzzy Multi-criteria Decision-making (___) Decision Tree Analysis
6	Time efficiency	(___) Bayesian-belief Networks (___) Fuzzy Multi-criteria Decision-making (___) Decision Tree Analysis
7	Usefulness in the current work	(___) Bayesian-belief Networks (___) Fuzzy Multi-criteria Decision-making (___) Decision Tree Analysis
8	Relevant to the current work	(___) Bayesian-belief Networks (___) Fuzzy Multi-criteria Decision-making (___) Decision Tree Analysis

	Acceptance of the proposed DSS model					
		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
	Perceived of Usefulness					
1	Using the DSS would improve my performance	☐	☐	☐	☐	☐
2	Using the DSS at work would improve my productivity	☐	☐	☐	☐	☐
3	Using the DSS would enhance my effectiveness in my job	☐	☐	☐	☐	☐
4	I would find the DSS useful in my job	☐	☐	☐	☐	☐
	Perceived Easiness of use					
5	Learning to operate the DSS would be easy for me	☐	☐	☐	☐	☐
6	I would find it easy to get the DSS to do what I want to do	☐	☐	☐	☐	☐
7	It would be easy for me to become skilful in the use of the DSS	☐	☐	☐	☐	☐
8	I would find the DSS easy to use	☐	☐	☐	☐	☐
9	My interaction with the DSS is clear and understandable	☐	☐	☐	☐	☐
10	The DSS is flexible to interact with	☐	☐	☐	☐	☐

11	Using the DSS could help me get the most out of my time to decide to open data	☐	☐	☐	☐	☐
	Intention to use					
12	I have the intention to use the DSS to decide to open data	☐	☐	☐	☐	☐
13	I have the intention to use the DSS when it becomes available in my office/department	☐	☐	☐	☐	☐
14	I have the intention to use the DSS when necessary to provide decisions in open data	☐	☐	☐	☐	☐
15	I have the intention to use the DSS in my daily activity or work	☐	☐	☐	☐	☐
16	Do you have any other comments and/or suggestions? If so, would you please write to them?					

End of the Survey

Dear participant,

Thank you very much for participating in this survey!

Curriculum vitae

Luthfi was born on 11 March 1976 in Curup Bengkulu, Indonesia. He holds his bachelor's degree in Management of Informatics in 1999 from Bina Darma University, Palembang. After graduation from his first degree, he joined as a lecturer in the field of informatics engineering at the Bina Darma University for more than a decade. In 2003, he obtained a scholarship from Asian Development Bank to continue his study at the master level, and he received his Master of Computer Science from Gadjah Mada University, Yogyakarta, in 2005. During his career in academia, Luthfi had several administrative positions in the Faculty of Computer Science at Bina Darma University, such as the head of the Informatics Department for five years (2007-2012) and the caretaker of Dean Faculty of Computer Science for several months. In September 2012, Luthfi decided to migrate to Yogyakarta, and he dedicated his new academic career at the Universitas Islam Indonesia at the Department of Informatics.

On 1 January 2017, Luthfi started his PhD journey at the Delft University of Technology, the Netherlands. He obtained a full scholarship for his PhD study from the Indonesia Endowment Fund for Education in collaborating with the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia (LPDP–BUDI LN). During his PhD research, he published and presented his research in a series of blindly peer-reviewed venues in various international conferences like IFIP eGOV-CeDEM-ePart, Digital Government Society (dg.o), I3E conferences and BMSD symposium. To understand the large-scale Engineering of Complex Systems based on Big Data analysis, Luthfi joined the IDEA League Doctoral School. Four different universities hosted sessions in TU Delft, RWTH Aachen Politecnico Milan, and the University of Göteborg.

Luthfi was also involved as a technical reviewer for scientific articles such as Government Information Quarterly, Information Polity, and Technology in Society, and Transforming Government. In order to disseminate his research concept, Luthfi introduced his decision-making support systems for opening data to several government domains in the Netherlands at the Unit Data Analytics and Big Data, Customs National Office in Rotterdam, and Department of Knowledge & Advice, National Archive in The Hague.

List of publications

[1] **Luthfi, A**., & Janssen, M. (2017). *A Conceptual Model of Decision-making Support for Opening Data*. Paper presented at the 7th International Conference E-Democracy, pp. 95-105, 14-15 December 2017, Athens, Greece. ISBN: 978-3-319-71116-4, ISSN: 1865-0929. https://doi.org/10.1007/978-3-319-71117-1_7.

[2] Yin, Y., Zuiderwijk, A., Van Ossenbruggen, J., Jeffery, K., **Luthfi, A**., Janssen, M. (2018). *How to improve policymaking using open data in Virtual Research Environments? An interactive workshop discussing privacy, security and trust strategies*. Paper presented at 19th Annual International Conference on Digital Government Research (DG.O), 30 May–1 June 2018, Delft, Netherlands. ISBN: 978-1-4503-6526-0. https://doi.org/10.1145/3209281.3209307.

[3] **Luthfi, A**., Janssen, M., & Crompvoets, J. (2018). *A Causal Explanatory Model of Bayesian-belief Networks for Analysing the Risks of Opening Data*. Paper presented at the 8th International Symposium (BMSD), pp. 289-297, 2–4 July 2018, Vienna, Austria. ISBN: 978-3-319-94213-1, ISSN: 1865-1348. https://doi.org/10.1007/978-3-319-94214-8_20.

[4] **Luthfi, A**., Janssen, M., & Crompvoets, J. (2018). A *Framework for Analysing How Governments Open Their Data: Institution, Technology, and Process Aspects Influencing Decision-Making*. Paper presented at the EGOV-CeDEM-ePart 2018, pp. 163-171, 3–5 September 2018, Donau-Universität Krems, Austria. ISBN: 978-3-903150-22-5, ISSN: 2524-1400.

[5] **Luthfi, A**. (2018). *Decision Support for Weighing the Potential Risks and Merits of Opening Data*. Paper presented at the EGOV-CeDEM-ePart 2018, pp. 305-312, 3–5 September 2018, Donau-Universität Krems, Austria. ISBN: 978-3-903150-22-5, ISSN: 2524-1400.

[6] **Luthfi, A**., Rehena, Z., Janssen, M., & Crompvoets, J. (2018). *A Fuzzy Multi-criteria Decision Making Approach for Analysing the Risks and Benefits of Opening Data*. Paper presented at the 17th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2018, pp. 397-412, 30 October-1 November 2018, Kuwait City, Kuwait. 978-3-030-02130-6, ISSN: 0302-9743. https://doi.org/10.1007/978-3-030-02131-3_36.

[7] **Luthfi, A**., & Janssen, M. (2019). *Open Data for Evidence-based Decision-making: Data-driven Government Resulting in Uncertainty and Polarization*. International Journal on Advanced Science Engineering and Information Technology, *Volume 9, Issue* 3, pp. 1071-1078. ISSN: 2088-5334.

[8] **Luthfi, A**., & Janssen, M. (2019). *A Comparative Study of Methods for Deciding to Open Data*. Paper presented at the 9th International Symposium (BMSD), pp. 213-220, 1-3 July 2019, Lisbon, Portugal. ISBN: 978-3-030-24853-6, ISSN: 1865-1348. https://doi.org/10.1007/978-3-030-24854-3_14.

[9] **Luthfi, A**., Janssen, M., & Crompvoets, J. (2019). *Decision Tree Analysis for Estimating the Costs and Benefits of Disclosing Data*. Paper presented at the 18th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2019, pp.

205-217, 18-20 Setember 2019, Trondheim, Norway. ISBN: 978-3-030-29373-4, ISSN: 0302-9743. https://doi.org/10.1007/978-3-030-29374-1_17.
[10] **Luthfi, A**., Janssen, M., & Crompvoets, J. (2020). *Stakeholder Tensions in Decision-making for Opening Government Data*. Paper presented at 10th International Symposium (BMSD), pp. 331-340, 6-8 July 2020, Berlin, Germany. ISBN: 978-3-030-52305-3, ISSN: 1865-1348. https://doi.org/10.1007/978-3-030-52306-0_23.
[11] **Luthfi, A**., Rukanova, B., Molenhuis, Janssen, M., & Tan, Y.H. (2020). *Bayesian-belief Networks for Supporting Decision-making of the Opening Data by the Customs.* Paper presented at the EGOV-CeDEM-ePart 2020, pp. 51-58, 31 August-2 September 2020, Linköping University, Sweden.
[12] **Luthfi, A**., & Janssen, M. (2021). *A Stakeholders Taxonomy for Opening Government Data Decision-making*. Paper presented at 11th International Symposium (BMSD), pp. 384-391, 5-7 July 2021, Sofia, Bulgaria. ISBN: 978-3-030-79975-5, ISSN: 1865-1348. https://doi.org/10.1007/978-3-030-79976-2_26.

9 789463 842440

TU Delft