Measuring citizen preferences for the Dutch Education Open Data Policy: A Path towards Citizen-Informed Decision Making

Darli Ciang
Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands
Student Number: 4624211

Abstract

This paper discusses the way individuals in their role as citizens make trade-offs between open education data attributes. The Dutch government agencies lack of insight in the citizen preferences for open data policy attributes lead them to evaluate and develop their open data policy only from the data provider perspective. In order to address the problem, this research aims to identify the citizens preferences for the open education data policy using the citizens stated choice experiments. The experiment is based on Random Utility Maximization theory, the study infers the citizen preferences of open education data attributes based on their choices for several hypothetical open education data policy. The study combines 4 attributes with 3 attribute levels to create the hypothetical open education data policy: *mode of information presentation, number of free engaging hackathon events, number of free citizen data skill training events*, and *risk of your personal data exposed to the public*. The citizens significantly value *risk of your personal data exposed to the public* and *mode of information presentation*. Government agencies have limited option if it wants to extend the open education data implementation because citizens reluctance to compromise the data protection attribute. This research is a first attempt to extend citizen stated choice experiment approach for the valuation of citizen preferences in the context of open education data policy. It provides an alternative method for governments to evaluate and develop their open data policy alongside the commonly used government/data provider perspective

Keywords: open data, open government data, policy, discrete choice experiment, open education data

1. Introduction

In the recent years, the governments throughout the world adopt open data policy. Several countries spearheaded the initiatives such as the United States with Open Government directive and Digital Government Strategy under Obama administration (Obama, 2009, 2012) and the European Union with Directive 2013/37/EU about the reuse of public sector information and the European Commission's Open Data Strategy (European Commission, 2003, 2011). The policy aimed for transparency, participation, citizen-government collaboration, evidence-based policy making, administrative efficiency, stimulate innovation, and economic growth (European Commission, 2011; Obama, 2009).

The Open Public Data is defined as data that: are paid for from the public purse and generated during or for the provision of a public service, are available to the public, are free of copyright and other third-party rights, are machine-readable and preferably comply with open standards (not PDF but XML, CSV, etc.), and can be re-used without restriction in the form of cost, compulsory registration, etc. (Algemene Rekenkamer, 2014). The Ministry of Education, Culture and Science (OCW) generates much education data that fits the definition of open public data which are interesting for the Dutch citizens. The Dutch open education data is chosen because the domain have long experience in openness and transparency of public data. Starting from publishing school performance data in 1997 and implementing the project "Windows for Accountability" in 2012 (Hanne Obbink, 2012), before the nation-wide commitment for open government action plan in 2013.

Zuiderwijk & Janssen (2014) compared seven Dutch governmental policies and found several characteristics of the Dutch open data policy such as lack of systematic collaboration and 'jumping on the bandwagon' tendency. The government agencies are susceptible to mimic the other agencies that it deemed successful and follow their "best practice" regardless of their data context and the environment they are operating in. Practitioners tendency to mimic other initiatives might lead them to overlook the objectives, the context and the deliverence of societal values which are unique to the domain they operate (Zuiderwijk, Shinde, & Janssen, 2018).

Susha et al. (2015) concluded that the creation of model and benchmarks should be guided from the perspective of what is beneficial for open data end users since it is the primary goal of opening data. The policymakers tendency to mimic each other and settle for generic performance indicators (quantity and benchmark scores) show their lack of insight in the citizen preferences for open data policy attributes.

Problem 1: Lack of insight in the citizen preferences of open data policy attributes

In order to address the problem, this research aims to identify the citizens preferences for the open data policy through the citizens stated choice experiment (CSCE). The citizen stated choice experiment is a type of discrete choice experiment (DCE). Discrete choice experiment is a quantitative technique to elicit individual preferences (Mangham, Hanson, & McPake, 2009).

The citizen preferences and measured trade-off attributes are important components to understand the existing gap between open data policy objectives and the realized benefits. In the current process, policymaker measure the policy performance from the data provider perspective. This

research could lead to a new performance indicator based on the citizens needs that the policy accommodate and how the citizens perceived the fulfillment. Other than that, identifying the citizen preferences in the agenda settings phase will help policymaker to accurately allocate their resource according to the citizens needs and prevent futile implementation. Furthermore, policymaker can create a citizen-informed decision making when they deal with various policy alternatives.

Therefore, the ultimate aim of this research is to empirically measured citizen preferences of open data policy attributes specifically the Dutch open education data. The results are meant to identify the relative preferences of the open education data policy attributes from the citizen perspective and how policymaker can utilize it to create a suitable open education data policy.

The main research question for this study is:

What are the preferences of citizens for a Dutch open education data policy?

In order to answer the main research question, the following research questions are derived:

1. What is the policy context (policy objectives, organization, existing implementation) of Dutch open education data policy?
2. What are the possible trade-off attributes for the open data policy in the existing literature?
3. How does the identified trade-off attributes and policy context translate into the citizen stated choice experiment design?
4. What is the valuation of each trade-off attributes for the respondents in their role as a citizen?
5. Considering the citizen preferences results, what are the recommendations to policymakers creating the Dutch open education data policy?

Sub question 1 and 2 is discussed in the Section 3. Conceptual Framework, sub question 3 is discussed in the Section 4. Experiment Design, sub question 4 is discussed in the Section 5. Experiment Results and sub question 5 is discussed in Section 6. Discussion and Conclusion

2. Methodology

The citizen stated choice experiment is a variant of discrete choice experiment (DCE) where the respondents are asked to do the choice tasks from “citizen” perspective instead of “consumer” perspective. Study by Mouter & Chorus (2016) differentiate consumer and citizen perspective based on different budget constraint. The consumer preferences when the choice involve after tax income of the individual and citizen preferences when the choice are based on previously collected tax by the government. In the study, Mouter & Chorus (2016) empirically confirm the difference between individual valuation of time gained depending on their role as “consumer” or “citizen”. Further research by Mouter et al. (2017a, 2017b) extend the notion of citizen stated choice experiments for other non-market goods valuation in transport policy such as safety and spatial equality.

The identification of preferences from citizen perspectives is suitable for open data policy because the provision of open data is fully-funded by the government. The Open Public Data is defined as data that: are paid for from the public purse and generated during or for the provision of a public

service, are available to the public, are free of copyright and other third-party rights, are machine-readable and preferably comply with open standards (not PDF but XML, CSV, etc.), and can be re-used without restriction in the form of cost, compulsory registration, etc. (Algemene Rekenkamer, 2014).

2.1. Citizen Stated Choice Experiment (CSCE)

There are several phases in designing choice experiments which are summarized in Table 1.

Table 1 Designing Discrete Choice Experiments

Phase (based on (Mangham, Hanson, & McPake, 2009))	Adaptation to this research	Relevant section
Establishing attributes	• Literature review of existing open data policy assessment study • Desk research on existing policy documents	Conceptual framework
Assigning attribute levels	• Desk research on existing policy documents	
Designing the choice sets	• Create balanced and orthogonal survey design	Experiment design
Generating, pre-testing, and distribute the questionnaire	• Pilot test questionnaire • Revised the questionnaire based on the input from pilot test • Distribute the final questionnaire	
Analyze DCE data	• Create a statistical model to analyze questionnaire results. • Explain the result of statistical model and its implication for policymaker	Experiment results

2.2. Random Utility Maximization Theory

The analysis of DCE data are based on the combination of two theory: 1) Lancaster's characteristic demand theory (Lancaster, 1966), and 2) Random Utility Theory (McFadden, 1974). The characteristic demand theory describes that consumer derived utility from the characteristics of goods rather than the consumption of the goods itself. This approach allows us to infer individual preferences based on their choice of characteristics (attributes) presented in the options.

The random utility theory allow researcher to analyze the utility derived from the goods characteristics. The amount of utility is represented by a relative and abstract numerical value, while choices are the only observable indicator of utility. Individuals expressed their preference from the amount of utility that they perceived, satisfaction when the specific attributes provide a positive utility and dissatisfaction for a negative utility. Other than that, the analysis is conducted on the basis that every individual is rationally maximizing utility who chooses an alternative that gives the largest

relative utility. The utility function can use linear or non-linear parameter. In its simplest form, the utility function can be defined as a linear expression in which each attribute is weighted by a unique parameter to account for that attribute's marginal utility (Mangham et al., 2009).

Key-elements of RUM-choice model:

- i, j = alternatives in the choice sets (i is alternative 1 and j is alternative 2)
- m = attributes (e.g. cost, time)
- X = attribute values from observation
- β = parameters to be estimated
- ε = randomness (all the unobserved determinants of the utility)

The systematic utility (V_i) is the utility that can be related to observed factors (e.g. cost, time, age, income level) which can be represented in the form of:

$$V_i = \sum_m \beta_m \cdot X_{im}$$

The total utility of an alternative can be represented through this equation:

$$U_i = V_i + \varepsilon_i = \sum_m \beta_m \cdot X_{im} + \varepsilon_i$$

An alternative is chosen if its total utility is the largest. Therefore, alternative i will be chosen over alternative j if it fulfills this condition:

$$\sum_m \beta_m \cdot X_{im} + \varepsilon_i > \sum_m \beta_m \cdot X_{jm} + \varepsilon_j, \forall j \neq i$$

Since the total utility (U_i) is the combination of systematic utility (V_i) and error term (ε_i). There will be a situation where an individual does not choose an alternative with the highest systematic utility due to the unobserved factors from error term. It implies that the prediction of choices is based on probability with an assumption (higher systematic utility → higher choice probability). The probability of alternative i is chosen over alternative j can be expressed as follow:

$$P(i) = P(V_i + \varepsilon_i > V_j + \varepsilon_j, \forall j \neq i)$$

It is important to note that the utility is not an absolute value. Therefore, what matters in the choice situation between alternative i and j is the utility differences of alternative i relative to alternative j. The probability equation can be rewritten as:

$$P(i) = P(U_i - U_j > 0, \forall j \neq i)$$

In this research, the multinomial logit model (MNL) is used to explicitly estimate the β which is the parameter that determine the individual preference/taste for a certain attribute/characteristic. The probability equation of choosing alternative i, if ε ~ EV Type 1 with variance $\pi^2/6$ in MNL model can be written as follow:

$$P(i) = P(V_i + \varepsilon_i > V_j + \varepsilon_j, \forall j \neq i) = \frac{e^{V_i}}{\sum_{j=1\ldots J} e^{V_j}} = \frac{e^{\sum_m \beta_m X_{im}}}{\sum_{j=1\ldots J} e^{\sum_m \beta_m X_{jm}}}$$

(Note: in the denominator, *J* denotes choice set size. *j* runs from 1 to , and includes *i*)

Estimating β implies inferring the importance of the attribute (e.g. cost) relative to other observed attribute (e.g. time) and relative to unobserved factors ('randomness/error term').

3. Conceptual Framework

3.1. Literature review on open data policy assessment study

The literature review is conducted through SCOPUS using the terms 'open government data', 'preference'. The terms 'measurement', 'assessment' and 'evaluation' are chosen to extend the scope of literature because using 'preference' term result in limited number of literature (24 journal papers). In the measurement, assessment, and evaluation study, the open data policy is scrutinized from different perspectives and aspects which are suitable to identify open data policy attributes.

The source type is limited to journal with the topic of social science and publication year between 2013-2018. The topic is limited to social science because there is similar study in computer science that focus on the technical side of open government data. For the purpose of identifying open data policy attributes a socio-technical perspective is needed, thus the social science is chosen as the subject area over the more technical computer science area.

Other than that, the literature are also found through snowballing method by looking at previous and subsequent study that cite the key literature such as Charalabidis, Alexopoulos, & Loukis' (2016) study about a taxonomy for OGD research.

Search terms:
(open AND government AND data AND (measurement OR assessment OR evaluation OR preference)) AND (LIMIT-TO (SRCTYPE , "j ") OR LIMIT-TO (SRCTYPE , "p ")) AND (LIMIT-TO (SUBJAREA , "SOCI ")) AND (LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013))

The initial search result in a total of 168 papers. A quick scan of the abstracts is performed on all papers to decide if they are relevant for this study. This step reduces the total number of papers to 44. Lastly, content analysis is performed, focusing on the aspect of the open data policy discussed and the perspectives of the study. The summary of literature can be found in Appendix A.

3.1.1. The tension between "stewardship" and "usefulness"

Government as the implementation agent of OGD program is faced with the inherent tension between the stewardship and usefulness principles in managing the information (Dawes, 2010). The stewardship principle focuses on the data provisioning dimension which address the issue of data confidentiality, information quality, information and system security, data management, and maintenance of data assets. On the other hand, the usefulness principle aims to foster the utilization of data to generate social and economic benefits which lead to strategies that improve public access to government information, stimulate public-private information partnerships, and innovative application of data.

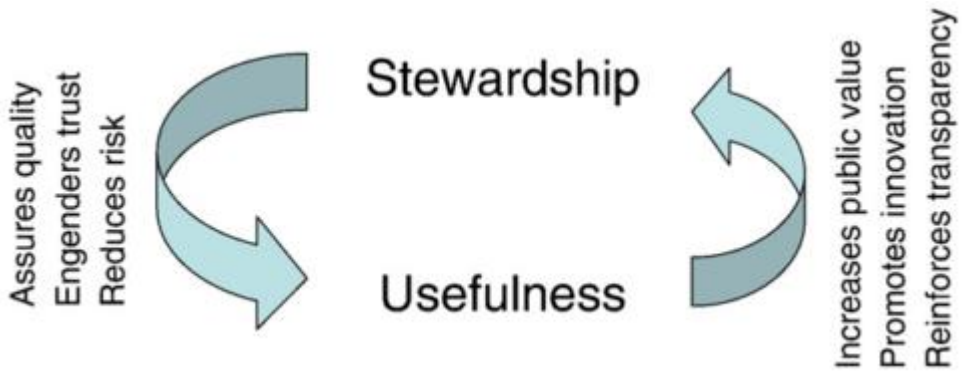

Figure 1 Conceptual model of information-based transparency principles adapted from (Dawes, 2010)

Further study by Reggi & Ricci (2011) assess 434 beneficiaries of EU Structural Funds and found that the open data strategy of those beneficiaries diverge into two clusters resembling the tension between stewardship and usefulness principles. "User centered" cluster focuses on the usefulness principle by providing data visualization and searching features, while "Re-user centered" cluster apply the stewardship principle by concentrating on data quality and validity. Lee & Kwak (2012) also differentiate data-related and participation/collaboration-related capabilities/processes in their 5 stage Open Government Maturity Model (OGMM), with the early stage focus on data capabilities and later stage on the participation/collaboration capabilities.

This tension is reflected in the existing benchmark and assessment studies where each study approach open data policy from diverse perspectives such as: the degree of dataset reusability, data quality, and other data provisioning attributes (Petychakis, Vasileiou, Georgis, Mouzakitis, & Psarras, 2014; Tim Berners-Lee, n.d.; Vetrò et al., 2016), the open data portal content and features (Afful-Dadzie & Afful-Dadzie, 2017; Lourenço, 2015; Thorsby, Stowers, Wolslegel, & Tumbuan, 2017; Zuiderwijk-van Eijk & Janssen, 2015), the user perspectives on open data usability (Weerakkody, Irani, Kapoor, Sivarajah, & Dwivedi, 2017), and the holistic approach which assess the open data program as an ecosystem (Ubaldi, 2013; Welle Donker & van Loenen, 2017)

3.1.2. The different perspectives in assessment study

OGD initiatives need to address challenges from different aspects (policy, legal, economic, organizational, technical, and cultural) in order to create an ecosystem that enable value creation (Ubaldi, 2013). The multi perspective nature of OGD initiatives are also reflected in the aspect that are being measured by different assessment studies and the perspective they took. Appendix A presents the summary of measured aspects and perspective taken by each study.

Open data portal assessment

Most of the studies assess the open data portal since it is the most common implementation of OGD initiatives. Even though the studies have an open data portal as the object of assessment, there are no standardization on which aspects to measure. For example, one study assesses the open data portal based on the *features* available in the portal (Sayogo, Pardo, & Cook, 2014), another study assesses the *characteristics of the data* published in the portal (Lourenço, 2015), even there is a study that assess the open data portal by looking at the *policy framework* and *socio-technical aspects* of the city where the open data portal is implemented (Chatfield & Reddick, 2017).

Sayogo, Pardo, & Cook (2014) assess open data portal based on: data content, data manipulation capability and participatory and engagement capability. Thorsby et al. (2017) compare cities open data portal in America based on its features and content diversity. The study categorizes features into content, help, policy, and results; each feature has different unit of measurement such as: can users search through datasets (help feature) or is there a call for action or invitation for citizens to use the data (results feature). The content, help, and result feature are comparable to data manipulation and engagement capability from Sayogo, Pardo, & Cook (2014). Chatfield & Reddick (2017) examines 20 local governments open data portal in Australia with the criteria: open data provision, data format variety, open data policy intensity, and entrepreneurial data services.

On the other hand, there are studies that specifically investigate the characteristic of datasets provided in the open data portal. Lourenço (2015) measured the open data portal based on desired characteristic of data disclosure which are: quality, completeness, access and visibility, usability and comprehensibility, timeliness, value and usefulness, granularity, comparability and compare nation-level open data portal using that criteria. Another study by Vetrò et al. (2016) analyze the data quality of open data portal using the established data quality metrics such as: completeness, accuracy, traceability, currentness, expiration, compliance, and understandability. Both studies investigate the datasets characteristic from data provider side. Afful-Dadzie & Afful-Dadzie (2017) use data-related metrics (data quality, data format, metadata, data availability, data integrity) in 5 African countries open data portal to inquire journalists attributes preferences for the portal. They found that the respondents chose metadata as the most important attributes with the relative importance weight of 28.82%, followed by data format (23.3%) and data quality (20.34%).

Zuiderwijk-van Eijk & Janssen (2015) create a quasi-experiment to measure the effect of participation mechanism and data quality indicators in the open data portal, the participants are assigned into the control and treatment group. The control group are asked to use the prototype open data portal which include the features such as discussion messages, social media sharing, linking items related to a dataset, wiki descriptions and discussions, and data quality ratings and reviews. The study suggested that participation mechanisms and quality indicators add value and improve the use of OGD portal.

The abovementioned studies analyze open data portal and there are recurring aspects from those studies such as: data manipulation capability, engagement capability, availability of open data policy, and non-technical features (promotion of open data, user's skills development, engagement events).

Socio-technical assessment

Next stream of studies focusses on a more comprehensive approach of assessing OGD by considering the socio-technical aspects of OGD. Study by Ubaldi (2013) provides an analytical framework and metrics of measurement on several dimension consists of: policies and law, technical, data governance, organizational, communication and interaction, political priorities, impact, and data-related metric such as availability, quality, uptake, re-use.

In the context of Dutch open data policy, two studies that apply socio-technical perspective are found. First, Zuiderwijk & Janssen (2014) creates a framework for comparing OGD implementation in seven Dutch government organizations. The study analyzes several aspects such as policy environment and context, policy content, performance indicators, and public values. The analysis is conducted from the data provider side (government) which can be seen from the information that

are being measured by each aspect. Examples of information being compared for the policy environment and context: level of government organization, resource allocation, legislation, socio-political context, culture of the institutions. The policy content aspect provides the specification of the OGD such as: target groups of open data, policy strategy and principles on publishing data, technical standards and formats of open data. The study found that most of the policies investigated focus on internal challenges to publish the data (privacy protection, confidentiality, data misuse and misinterpretation, embargo periods, data quality, data completeness) and less concern on the usability of the data (how it can be used to create the desired public values.

Second, Welle Donker & van Loenen (2017) examines the Dutch open data ecosystems from two aspects: data supply indicators (known, attainable, usable) and data governance indicators (vision, leadership, self-organizing ability, financing, open data stimulation, supply-user communication, G2G communication). The data supply indicators investigate whether the dataset is searchable and can be found for use (known), accessible from a financial, legal, and practical aspect (attainable), and (usable) in terms of having complete metadata, documentation, and up-to-date. The study not only analyze open government data from the government perspective but also ask the infomediaries (users who developed services using open data) about the open data governance. The study found that infomediaries criticize the existing data governance model where government waiting for the creation of "killer app" and organize hackathon with temporarily available datasets. The infomediaries would prefers government to develop a sustainable open data business model; being a launching customer and commission the infomediaries to develop open data tools and applications.

Dawes, Vidiasova, & Parkhimovich (2016) assess OGD programs in New York and St. Petersburg within the dimensions of settings, motivation, policy and strategy, data publication and use, feedback and communication, benefits, and advocacy and interaction among stakeholders. The dimensions used are comparable with the abovementioned study; for example, settings, motivation, and policy & strategy is similar with Zuiderwijk & Janssen (2014) policy environment and context and policy content. The data publication and use are comparable with data supply indicators (known, attainable, usable) from Welle Donker & van Loenen (2017) study.

Studies in the socio-technical streams mainly focus on the policy context and strategy of the government. However, it also discusses the data-related attributes and participation & engagement related attributes extensively. Even though it uses different terminology, for example data supply/policy content/data publication and use for the data-related attributes; communication and interaction/feedback and communication/open data stimulation/supply-user communication for the participation & engagement related attributes.

Citizen assessment

Recent studies start to investigate open government data from the citizen perspectives. Weerakkody et al., (2017) measures the citizen intention to use open data using the modified and extended Diffusion of Innovation (DOI) model with predictors: relative advantage, compatibility, observability, and security risk. The study found that relative advantage, compatibility, and observability are statistically significant on predicting citizens intention to use open data. The security risk had no significant effect on citizen intention to use open data. The study suggests that most citizens have no

concerns about trusting public sector open data and do not perceived a significant security risk in the open data.

Zuiderwijk et al. (2018) investigate the attainment of OGD objectives based on the delivered benefits which are categorized into operational, technical, economic, and societal benefits. The study shows that the most delivered benefits are operational and technical benefits, followed by economic benefits, and societal benefits. The study also concludes that there is a mismatch between open data objectives and the delivered benefits. Achievement of the benefits are not significantly related to the presence of objective related to the delivery of the benefits.

Safarov, Meijer, & Grimmelikhuijsen (2017) conduct a systematic literature review on the utilization of open government data and identify the conditions for utilization. In the study they review 101 studies and found two categories of condition for utilization which are technical and social condition. Technical conditions refer to the feature of OGD such as the data quality, data availability, and infrastructure to enable OGD; Social conditions refer to the institutional context (policy, legislation, organization) and the skills of users. The study also found distinction between users, direct users who use the OGD themselves and indirect users who use the data/services processed by intermediaries.

In Table 2 the identified assessment attributes from existing studies are presented. Reviewed studies repeatedly use variance of data-related and participation & engagement related aspects in their analysis. Therefore, in this study the trade-off attributes are categorized into data-related attributes and participation & engagement attributes as shown in Table 2**.** The categories also reflect the tension between data stewardship and usefulness principles discussed in section 3.1.1. Other than that, there are attributes specifically related with the usability, communication, and interaction features of the open data portal.

In the identification process, this research only selects attributes which can be experienced directly by the citizens. Therefore, aspects that discuss the internal arrangement of the data provider are excluded. For example, intergovernmental agency communication, organization restructuring, political priority.

Table 2 Open Data Policy attributes from existing assessment study

Category	Attributes	Study
Data-related attributes	Data Availability (number of datasets, API)	(Afful-Dadzie & Afful-Dadzie, 2017; Petychakis et al., 2014; Safarov et al., 2017; Sayogo et al., 2014; Thorsby et al., 2017; Ubaldi, 2013; Welle Donker & van Loenen, 2017)
	Data Quality (accuracy, consistency, update timeliness, completeness)	(Afful-Dadzie & Afful-Dadzie, 2017;

		Petychakis et al., 2014; Safarov et al., 2017; Thorsby et al., 2017; Ubaldi, 2013; Vetrò et al., 2016; Welle Donker & van Loenen, 2017; Zuiderwijk-van Eijk & Janssen, 2015)
	Data Discoverability (advanced search tools on portal, metadata)	(Afful-Dadzie & Afful-Dadzie, 2017; Attard, Orlandi, Scerri, & Auer, 2015; Petychakis et al., 2014; Thorsby et al., 2017; Welle Donker & van Loenen, 2017)
	Data Protection	(Attard et al., 2015; Safarov et al., 2017; Weerakkody et al., 2017)
Portal-related attributes	Communication and Interaction in Open Data Portal	(Petychakis et al., 2014; Safarov et al., 2017; Sayogo et al., 2014; Thorsby et al., 2017; Titah, 2017; Ubaldi, 2013; Zuiderwijk-van Eijk & Janssen, 2015)
	Open Data Portal ease of use	(Safarov et al., 2017; Thorsby et al., 2017; Titah, 2017; Weerakkody et al., 2017)
Participation & engagement related	Public Awareness	(Attard et al., 2015; Thorsby et al., 2017; Weerakkody et al., 2017; Welle Donker & van Loenen, 2017)
	Public Participation (citizen involvement in promoting, using, and discussion about open data)	(Attard et al., 2015; Titah, 2017; Welle Donker & van Loenen, 2017)

	Motivation (competition, public-private partnership)	(Attard et al., 2015; Weerakkody et al., 2017; Welle Donker & van Loenen, 2017)
	Development of required skills and expertise to use Open Data	(Safarov et al., 2017; Welle Donker & van Loenen, 2017)
	Compatibility (the provided open data suit the needs of the citizen)	(Weerakkody et al., 2017; Welle Donker & van Loenen, 2017)
	Data Reusability (number of application created, number of new services from open data)	(Sayogo et al., 2014; Thorsby et al., 2017; Ubaldi, 2013)

From the literature review, three common categories of open data policy attributes are identified: data-related attributes, portal-related attributes, and participation & engagement related attributes. The data-related attributes consist of data availability, data quality, data discoverability, and data protection. The portal-related attributes are communication and interaction features, open data portal ease of use. Finally, the participation & engagement related attributes are public awareness, public participation, motivation, development of required skills and expertise, compatibility of the data provided with the needs, and data reusability.

3.2. Open education data policy in the Netherlands

Ministry of Education, Culture, and Science (OCW) publish its open data in several portal such as duo.nl, onderwijsinspectie.nl, and onderwijsincijfers.nl. All these data are also registered in the national open data portal called data.overheid.nl. Each of the portal have different type of data generated by the respective government agency (DUO, Education Inspection Agency, and OCW). The data is presented in the form of original source and static/dynamic figures. Other than that, there are limited functional applications resulted from the open datasets such as scholenopdekaart.nl and studiekeuze123.nl.

Furthermore, OCW arrange annual event for open education data which brings parents, students, teachers and school management together to discuss the possible application of open data. It is arranged in November 2016 with the theme “Education Data under scrutiny”. In this event, the participants came with several ideas to utilize open data which lead to one question as a use case “How can I make a good secondary school choice based on my values?” (Rijksoverheid, 2016). Afterwards, the event hosted a hackathon to create application prototype that answer the use case. In 2018, OCW and municipality of Amsterdam organise a hackathon “Hack de Valse Start” to address the question of inequality opportunities in the education (openstate.eu, 2018). This hackathon aims to gain insight on inequal opportunities by combining education and municipality open data provided by DUO and Central Bureau of Statistics.

In 25th May of 2018, the General Data Protection Regulation (GDPR) is formally applied in the Netherlands. The introduction of GDPR reinforce the existing barrier faced by government agencies in opening their data (risk-averse culture and limited resource to handle the data publishing process). The risk of opening data is increased because there is a hefty fine in case of data breaches (as high as €20 million or €10 million according to the bill).

In order to comply with the data protection specification of the GDPR, sizeable resources are required (both personnel and monetary) which will put a pressure on their budget for other functions. OCW hires two Data Protection Officers, one at DUO and one at the board department. A specific FG at DUO was chosen because of the large amount of personal data at DUO and the need to exercise adequate supervision at a short distance (OCW, 2017). The Data Protection Officer is in charge of Data Protection Impact Assessments (DPIA), mapping the privacy risks of a data processing system in advance and take measures to reduce the risks.

For example, DUO requires the amount of €12 million in 2018, increasing to €27 million in 2022 to implement the changes required by GDPR; government concludes that with the existing problems in OCW budget no room for this expenditure within the 2018 budget (OCW, 2017).

The existing open data policy focus on the data stewardship capability to ensure the supply of open data. However, there are limited functional applications resulted from the open datasets such as scholenopdekaart.nl and studiekeuze123.nl. The introduction of GDPR also create another pressure for the government agencies in charge of open education data (OCW, DUO, and Education Inspection Agency).

3.3. Discussion

Three categories of potential open data policy attributes are identified in Chapter 3 which are *data-related attributes*, *portal-related attributes*, and *participation & engagement attributes*. Based on the policy context exploration, there are significant implementation of data-related attributes and participation & engagement attributes within Dutch open education data policy.

The Ministry of OCW provides information in diverse forms such as: raw data in the respective open data portals (DUO, OCW, Education Inspection Agency), static and interactive figures (OCW and VSNU portals), and creating services from open education data (scholenopdekaart.nl and studiekeuze123.nl). Other than that, several participation & engagement events are organized such as data exploration event “Education Data under scrutiny” and hackathon “Hack de Valse Start”. There is no specific portal-related attributes implementation in the OCW open education data policy, all the data are simply hosted in each agency open data portal without any additional features for the users to interact with the portal.

However, one aspect of data-related attributes is growing in importance based on the policy context. The increasing importance of data protection attribute is influenced by the passing of General Data Protection Regulation (GDPR) in 25th May of 2018.

The introduction of GDPR reinforce the existing barrier faced by government agencies in opening their data (risk-averse culture and limited resource to handle the data publishing process). The risk of opening data is increased due to a hefty fine in case of data breaches. Sizable resources are required (both personnel and monetary) to fulfil the GDPR data protection specification. This condition put a

pressure on the already limited budget and personnel of government agencies in charge of open education data (DUO, OCW, Education Inspection Agency). The complexity of opening data is increased due to the additional requirement of Data Protection Impact Assessments (DPIA).

Based on the policy context exploration the three category of open education policy attributes is modified into data-related attributes, data protection attribute, and participation & engagement related attributes. The portal-related attributes are omitted because in the context of open education data policy there is only a basic open data portal implementation.

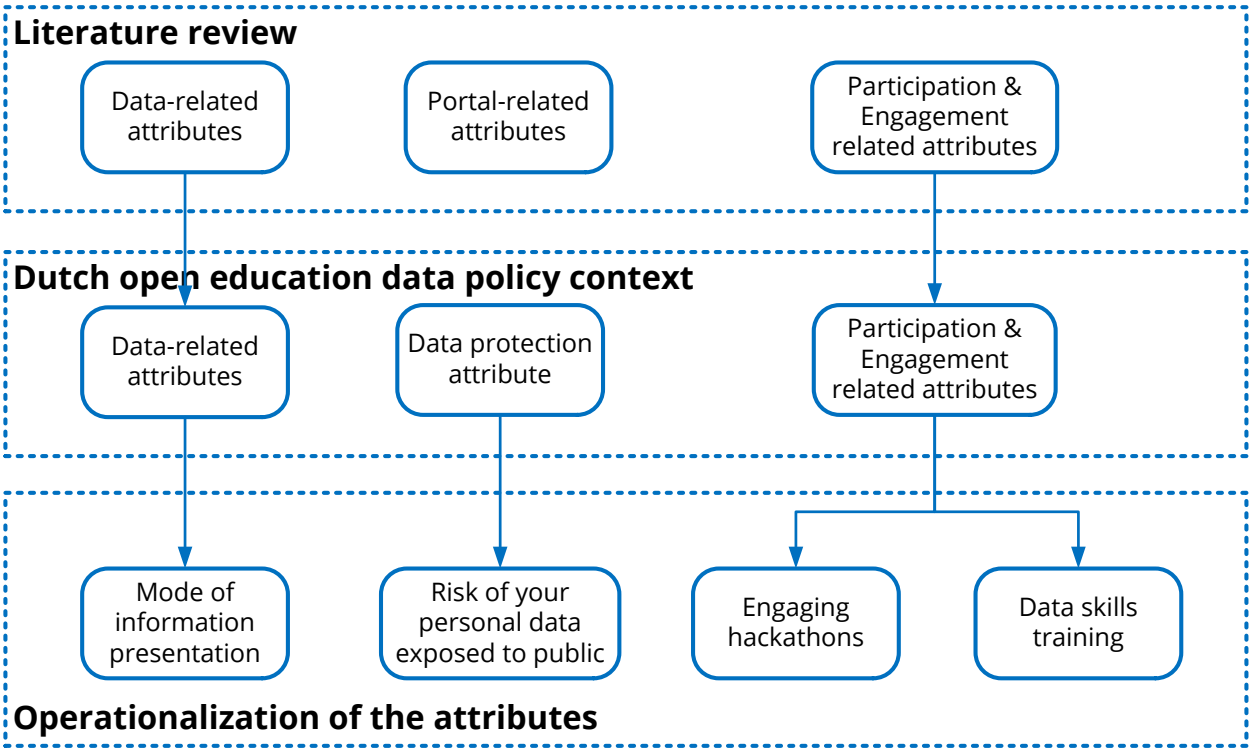

Figure 2 Conceptual framework of open education data policy attributes

4. Experiment Design

4.1. Attribute selection

The selection of attributes is based on several criteria:

- Expected influence on an individual choice (in this context the Dutch higher education students as the target respondents)
- Societal relevance of the factor (whether the attributes complement the Dutch open education data policy motivation for *education quality openness*)
- Measurability in the discrete choice experiment (whether the attributes have tangible unit of measurement and can be operationalized for the choice situations)

The three category of open education policy attributes is modified into data-related attributes, data protection attribute, and participation & engagement related attributes as shown in Figure 2.

For the data-related category, mode of information presentation is selected as the attribute. It is assumed that the provided data meets the data quality standard (accuracy, consistency, update timeliness, completeness), complete metadata, and accessible in standard format.

The participation & engagement related attribute is the umbrella term for diverse type of activities to stimulate the public participation such as: public training to increase the citizen data proficiency, hackathon to create new services, data exploration event to identify public data needs, support for monthly meeting of civic innovators. Therefore, two attributes are defined for the participation & engagement related attribute which are: engaging hackathons and data skills training

Finally, the risk of respondent personal data exposed to the public is selected for the data protection category

The discrete choice experiment considers labelled and unlabelled alternatives. Labelled alternatives are used when the labels represent characteristics not varied in experiment. For example, DCE for mode of transportation have label specific characteristic such as: car, train, plane. Each alternative has specific characteristics that are not varied or there are alternative specific attributes e.g. different range for travel time, parking fee for car.

In this experiment the unlabelled alternatives are used because both alternative use the same generic attributes and there is no label specific characteristic, the alternatives are simply called Policy A and Policy B.

Table 3 Overview of level of measurement and unit of measure

Attributes	Level of Measurement	Unit of Measurement
Access to the information	Nominal	Mode of information presentation
Participation & engagement activities	Ratio	Number of free engaging hackathon events Number of free citizen data skill training events
Data protection	Ratio	Number of data leak incidents

Table 4 Specification of parameters for pilot design

Attributes	Parameter
Mode of information presentation	β_1Data
Engaging hackathons	β_2Hackathon
Data skills training	β_3Training
Risk of your personal data exposed to public	β_4Privacy

4.2. Attribute levels

Access to the information

Access to the information attribute reflects different mode of information presentation that are currently implemented by OCW. Three different mode of information presentation are available: in original form (as similar as possible to the source) as the base value, static or interactive figures, and functional services (e.g. an application such as studiekeuze123.nl or scholenopdekaart.nl).

Participation & engagement events

In the survey the participation & engagement events are represented by two attributes: free engaging hackathon event and free citizen data skill training event. The reason behind it is to provide a more concrete attributes for the respondent to compare rather than a generic term of participation & engagement events. The term hackathon and data skill training can be specified in its aim and the benefits provided.

Therefore, the experiment chooses 1 free engaging hackathon event per 2 year and 1 free citizen data skill training event per year as the base value for participation & engagement events attribute. The attribute levels are scaled up to (1 event per year and 2 events per year) for free engaging hackathon events and (2 events per year and 3 events per year) for free citizen data skill training events.

Data protection

OCW annual report in 2017 record that there are 47 cases of data breaches reported within DUO and 3 cases of data breaches in OCW (OCW, 2018). 20 cases of the data breaches in DUO have been reported to the Dutch Data Protection Authority according to the regulations. There is no information about when the data breach happens, what type of data are compromised, and from what channel the data breach happens.

The existing open education data is highly deanonymized and only contain the aggregate information which cannot be traced to the individual. However, if there is a need for fine-grained data for a certain use case such as a hackathon that require the social background information of the students, the data will be more susceptible to be compromised. Therefore, the experiment chooses 1 incident per year as the base value followed by 1 incident per quarter and 1 incident per month as the range for the number of data leak incidents.

Table 5 Overview of attribute level and value

Category	Attributes	Value
Data-related attribute	Mode of information presentation	• in original form (as similar as possible to the source) • as static or interactive figures • as a service (e.g. an application such as studiekeuze123.nl or scholenopdekaart.nl)
Participation & engagement related attribute	Number of free engaging hackathon events	• 1 every 2-years • 1 per year • 2 per year
	Number of free citizen data skill training events	• 1 per year • 2 per year • 3 per year

Data protection attribute	risk of your personal education data exposed to the public	• 1 incident per year • 1 incident every 3-months • 1 incident per month

4.3. Pilot Survey

The pilot survey is the first phase of the survey design process to test the survey with small number of respondents and collect feedback on the survey length and understandability. The feedback from respondents can be used to adjust attribute levels. The following paragraphs elaborate on the different design steps to design a pilot survey. These steps are:

1. Model specification
   The pilot study model contains two unlabelled alternatives, labelled attributes and no alternative specific constant (ASC). The utility functions for the two alternatives are shown in the equation:

   $$U(alt1, alt2) = B\_Wdata * Xdata + B\_Whackathon * Xhackathon + B\_Wtraining * Xtraining + B\_Wprivacy * Xprivacy + \varepsilon$$

Variable	**Definition**
U(alt1, alt2)	Utility function for policy A and policy B
B_Wdata	Generic parameter for the attribute mode of information presentation
B_Whackathon	Generic parameter for the attribute free engaging hackathon events
B_Wtraining	Generic parameter for the attribute free engaging data skill training events
B_Wprivacy	Generic parameter for the attribute data protection
ε	Random error component

2. Generating experimental design
   A fractional factorial orthogonal design is selected to estimate the most reliable parameters with the lowest standard errors. Full factorial designs are not feasible because this leads to too many choice situations: 3^4 = 81. Therefore, a fractional factorial orthogonal design using basic plan 2 design is chosen, with three attributes in three levels and a total of 9 choice sets. The experiment is generated using Ngene software with a sequential construction of the alternatives. The complete list of choice situations can be found in Appendix B

3. Constructing the survey
   The online survey program SurveyGizmo is used to design the full pilot survey. The survey is constructed in English and distributed to Dutch respondents within the network of friend.

   The pilot survey consists of three different parts:
   - Leading questions about open education data policy
   - Choice situations
   - Perception and demographic questions

The final survey is improved based on the following feedbacks: 1) include the description for each attribute, 2) reduce the wordiness of choice situations, and 3) clearly define the extent of data leakage.

5. Experiment Results

5.1. Sampling procedure and descriptive result

The final survey is distributed among students who are currently attending Dutch higher education institution or recently graduated. The higher education students are targeted due to several reasons:

- Higher education students have relevant use case for the open education data which make them more likely to know about open data. (e.g. use open education data for courses, use the service to search for study programme)
- Higher education students have relevant skills to use open education data which make them more likely to be motivated on using open education data. (e.g. data analysis skill, programming skill)
- Higher education students are more likely to understand the term used in the survey with a proper explanation compared to other potential respondents (i.e. parents, primary/secondary school students).

The online survey is distributed through the network of friends and self-distributed in Delft University of Technology. The survey obtained 59 respondents from 18-30 June 2018. The summary of sample characteristics is presented in Table 6.

Table 6 Sample characteristics

Gender	Count	Percentage	
Male	47	79.66%	
Female	11	18.64%	
I do not want to specify	1	1.69%	
Age	**Count**	**Percentage**	
18 - 24	39	66.10%	
25 - 30	18	30.51%	
Above 30	2	3.39%	
Education	**Count**	**Percentage**	**Specialization**
HBO (hoger beroepsonderwijs)	6	10.17%	I do not want to specify = 6 Business & Economics = 1 Law = 1 Building engineering = 1 Educational studies = 1 Information Science = 1

WO (wetenschappelijk onderwijs)	53	89.83%	I do not want to specify = 6 Electrical engineering = 2 Engineering and Policy Analysis = 3 Complex Systems Engineering and Management = 9 Complex Systems Engineering and Management (ICT) = 2 Complex Systems Engineering and Management (B&S) = 1 Complex Systems Engineering and Management (Energy) = 2 Complex Systems Engineering and Management (T & L) = 1 Architecture = 1 Economics = 1 Civil Engineering = 4 Industrial Engineering and Management (IEM) = 1 Mechanical Engineering = 6 Technology, Policy and Management = 5 Chemical Engineering = 1 System and Control = 1 Clinical Technology = 2 Computer Science = 2 Microbiology = 1 Economics = 1 Design for Interaction = 1

The descriptive results show that majority of the respondents are familiar with open education data portal and the services created from open education data. 64% of the respondents have visited at least 1 open education data portal and 61% of the respondents have used at least 1 service created from open education data. However, only 7% of the respondents have attended open education data events.

5.2. Model specification

The survey result is modelled as the MNL (Multinomial Logit) model to estimate the relative values of open education data attributes for the respondents. The MNL model is suitable for the goal of this research which is to estimate the citizen preferences of open data policy attributes.

The model can be used to gain insight on the main effect of each attribute toward the citizen perceived utility. The MNL model on the probability of individual i choosing alternative q is shown in the following equation:

$$Piq = P\ (i\ |Cq) = \frac{e^{V_{iq}}}{\sum_{j \in Cq} e^{V_{iq}}}$$

Where:
Piq is the probability an individual i chooses alternative q
Viq is the utility of individual i to choose alternative q
Cq is the choice set of j alternatives for individual i

The MNL model parameters is specified in Table 7. The utility parameters for *"number of free engaging hackathon events"* (B_Whackathon), *"number of free citizen data skill training events"* (B_Wtraining), and *"risk of your personal data exposed to the public"* (B_Wprivacy) are estimated linearly. The attribute *"mode of information presentations"* (*B_Wdata*) is dummy coded where the levels represent the complexity of implementation. The dummy coding scheme is sketched in Table 8.

Table 7 MNL Model Parameters Specification

MNL Model Parameter Specification	
Variable	Parameter
B_Wdata_raw	*βdata_raw*
B_Wdata_figures	*βdata_figures*
B_Wdata_services	*βdata_services*
B_Whackathon	*βhackathon*
B_Wprivacy	*βprivacy*
B_Wtraining	*βtraining*

Table 8 Dummy coding for attribute "mode of information presentation"

	β_DATA_RAW	β_DATA_FIGURE	β_DATA_SERVICE
Level 2: Data as services	0	0	1
Level 1: Data as figures	0	1	0
Level 0: Data in original form	1	0	0

Hypotheses for the signs of the utility parameters are set up:

Hypothesis 1: negative estimate sign for the attribute "*risk of your personal data exposed to the public*"

Hypothesis 2: positive estimate signs for *"number of free engaging hackathon events"* and *"number of free citizen data skill training events"*

Hypothesis 3: positive estimate sign with non-linear utility for *"mode of information presentation"*.

For attribute "*risk of your personal data exposed to the public*", it is expected that the increasing data breach from 1 incident per year until 1 incident per month will result in a decrease in a respondent's utility for an alternative.

Increasing number of participatory & engagement events will result in an increase for the utility derived by respondents from an alternative. For *"mode of information presentation",* it is expected that the change from basic mode of information (data provided in original form) to the next attribute level (data provided as services) will increase respondent's utility for an alternative. The attribute levels are represented in ordinal values, hence the utility value derived from each attribute level cannot be estimated linearly.

5.3. Model estimates

Table 9 Model estimates without checking for linearity

Observations	531			
Individuals	59			
Rho-square	0.121			
Variable	Estimation	Standard Errors	t-test	p-value
βDATA	0.332	0.0935	3.55	0
βHACKATHON	0.0352	0.113	0.31	0.75
βPRIVACY	-0.702	0.0903	-7.78	0
βTRAINING	0.0748	0.0947	0.79	0.43

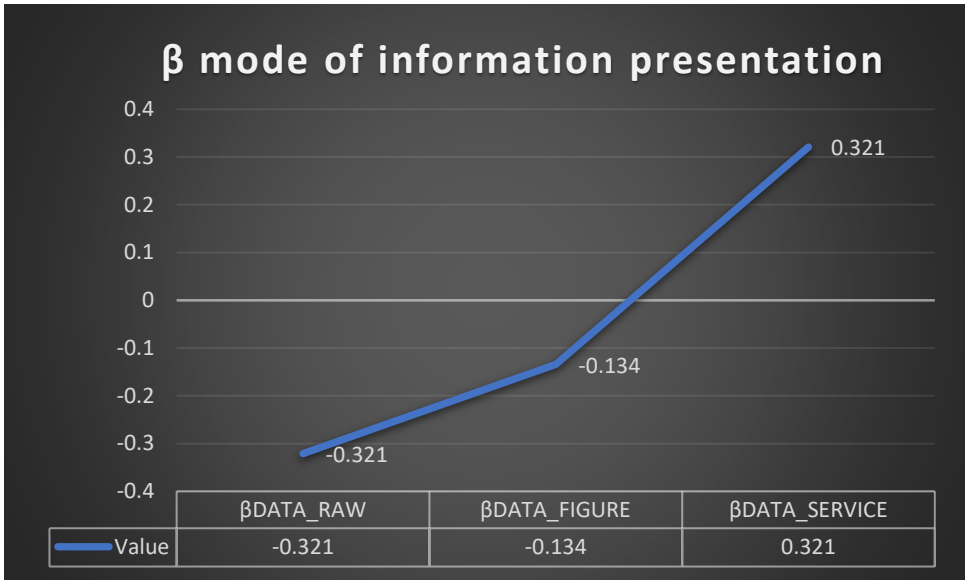

Figure 3 mode of information presentation utility

The model result in Table 9 shows that two of the most significant attributes are *"mode of information presentations"* and *"risk of your personal data exposed to the public"*. The "*risk of your personal data exposed to the public"* is statistically significant with an estimation parameter of -0.702 which means that an increase of incident frequency from 1 incident per year to 1 incident per quarter will reduce the utility of an alternative by 0.702. The *"mode of information presentations"* is a non-linear attribute as shown in Figure 3**Error! Reference source not found.**, significant improvement of the utility is identified when the data is presented as a service with 0.455 utility gain compared to data presented as a figure and 0.642 utility gain compared to data presented in an original form.

The attributes *"number of free citizen data skill training events"* and *"number of free engaging hackathon events"* are insignificant in the model. This result is congruent with the most and least important attributes from the descriptive results which shows that the respondents are consistent in their choices.

5.4. Goodness of fit

The McFadden's Rho-squared statistic is typically measured to evaluate the model fit. The Rho-squared expresses the level of uncertainty the model reduces, compared to a model with all zero estimations. The rho-square of 0.121 signifies that the estimated model is able to reduce the level of uncertainty by 12.10%, compared to a model with all zeros. Therefore, the MNL model's ability to predict citizen choices between the alternatives is arguable. However, the model is still suitable to identify statistically significant attributes.

6. Discussion and Conclusion

In this section, the implication of citizen preferences for open education data policy and the recommendations for policymaker will be discussed.

6.1. The assumptions and limitations of the study

Before discussing the implications of the results, there are several assumptions and limitations in this study. First, the target respondents for this study is limited to Dutch higher education students and the content of the survey is designed to fit their context. Therefore, the result of this study is based on the preferences of Dutch higher education students in their role as a citizen.

Second, the descriptive result shows that only 7% of the respondents have attended open education data events. It can be a reason for hypothetical bias (respondents choose attributes that are familiar for them). Replicating this study with a more balanced sample of respondents (who have experienced all the attributes presented in the questionnaire) will give a better insight on whether the respondents have a true strong preference for *"risk of your personal data exposed to the public"* and *"mode of information presentation"* and not from alternative explanations (e.g. misunderstanding, boredom, strategic behaviour).

Third, this research is the first attempt to empirically measures citizens preferences for open education data policy attributes and by no means set definitive valuation of trade-off attributes discussed in this study. I believe that the attributes estimation obtained in the study are reasonable and reflect Dutch higher education students' preference of open education data policy in their role as a citizen. However, further replication of the study with more diverse respondents are needed for conclusive valuation of attributes presented in this study. The result of this study should become the basis for further academic discussion and investigation.

6.1. The citizen preferences of open education data policy attributes

The result of citizen stated choice experiment shows citizens significant preference for *"mode of information presentations"* and *"risk of your personal data exposed to the public"*.

The *"risk of your personal data exposed to the public"* has the highest utility estimation of -0.702 which means that movement from the base attribute level of 1 incident per year to 1 incident per quarter will reduce the utility of an alternative by 0.702 for the citizen.

The *"mode of information presentations"* is a non-linear attribute with a slight difference of 0.187 utility estimation between the information presented in the original form and in the figure. There is a significant utility gain if the information is presented as a service compared to other attribute levels, 0.455 utility gain over information presented in the figures and 0.642 utility gain over information

presented in the original form. However, the gain is not enough to offset the utility reduction from *"risk of your personal data exposed to the public"* which can explain the dominant alternative in several choice situations.

The attributes *"number of free citizen data skill training events"* and *"number of free engaging hackathon events"* are insignificant in the model. This result is congruent with the most and least important attributes from the descriptive results which shows that the respondents are consistent in their choices. However, the descriptive result shows that only 7% of the respondents have attended open education data events. It might have been difficult for respondents to assess their preferences for participation and engagement events (hackathon/data skills training) if they have never attended one.

Citizens give significant negative response on any policy which compromise their personal data protection. The respondents do not want to trade their personal data protection with any type of improvement in the other attributes. It gives the government agency limited choices to improve the open data policy because the risk for opening data and compromise the data privacy is higher for them than the benefits that the other attribute can deliver.

However, given the description of open education data breach as follow: "*The open education data is anonymized. The personal data leakage happens when a person can be identified from the combination of multiple anonymous open datasets*". 58% of the respondents shows no concern about the possibility of data privacy breach. It seems in reality respondents have great trust on government to protect their privacy and the wording of choice situation exaggerate the possibility of data breach. The policymaker should consider this fact in the interpretation of this study and further investigation is needed to conclusively determine the utility of data protection attribute.

6.2. Recommendations for policymakers

Even though the existing policy suits the citizen preferences. There are several recommendations for government to improve the open education data policy:

1. Collaborate with infomediary to provide services for citizens
   The citizens derived significant utility from the information that is provided as a service compared to the other forms (original data and figures). However, the citizens lack of motivation to contribute for service creation will lead to bottleneck on the creation of new services based on open education data.

   It is important for the government to build partnership with infomediary users which use the raw open education data to create functional services for other citizens. The government can collect the requirements for new services during its annual event "OCW Kennisfestival" and consult with the respective stakeholders (education council, students, parents) afterwards for the detail specifications. After that, the government can commission the creation of the service to the infomediary users. In this process, the government can use monetary incentives to motivate the infomediary users for creating the service.

   Commissioning the service creations to infomediary users also enable the services to use the education data that are not publicly available. Government can provide those data directly to them and control the handling of the data. For example, studiekeuze123.nl have non-public

data from National Student Survey that it uses for measuring student satisfaction in the study program. The non-public data may have usable information that cannot be disclosed according to the privacy assessment model in the usual open education data. Other than that, citizens are more likely to trust and use services that are officially commissioned by the government.

The Ministry of Education, Culture and Science has enough experience in this schema of partnership with infomediary users as can be shown from the "Windows for Accountability" project which lead to the creation of scholenopdekaart.nl and studiekeuze123.nl.

The schema also produces the highest possible utility combination because government can provide the information in the form of services while protecting the citizens personal data internally. However, given the limited budget of the government there will be a trade-off between this recommendation and immediate needs to comply with the GDPR requirements.

2. Engage the citizens in a cost-efficient and subtle manner
   The two least significant attributes are *"number of free citizen data skill training events"* and *"number of free engaging hackathon events"*. Between these attributes, the citizens prefer *"number of free citizen data skill training events"* over the *"number of free engaging hackathon events"*. It can be interpreted that citizens prefer improvement of the data literacy of the general population rather than the one-time event such as a hackathon.

   It is expected because majority of the citizens are not interested to participate in the hackathon events, but they may perceive data skill training events as more beneficial for the general population.

   The data skill training can be implemented in different forms:
   - One of the respondents recommends creating online course that can be freely accessed by the citizens.

     > *"Would it not be far more convenient for a lot of people to create, for instance, an online learning environment for people to get acquainted with open data?"*

   - Another respondent recommends engaging the students from the early level of education. Government can embed the data literacy skills in the education curriculum as well.

     > *"Most people likely never heard of training events and many of the hackatons are also likely new to people. I reckon that marketing would be better if people are made enthusiastic at elementary schools and high schools rather than when they are mature already."*

6.3. Conclusion

Finally, the main question is addressed:

"What are the preferences of citizens for a Dutch open education data policy?

Based on the citizen stated choice experiment, the Dutch higher education students in their role as a citizen significantly valuate data protection attribute (*"risk of your personal data exposed to the public"*) and data-related attributes ("mode of information presentation").

Between three type of "mode of information presentation", citizens derive significant value if the data is presented as a service compared to data presented as a figure, and data presented in an original form. However, the value gained from the improvement in "mode of information presentation" is not enough to offset the loss of value in case of data breach.

Therefore, government agency has limited choices to improve the open data policy because the risk for opening data and compromise the data privacy is higher for them than the benefits that the other attribute can deliver.

However, the possibility of 'hypothetical bias' should be considered in the interpretation of the result. In the survey, open education data breach is described as follow: "The open education data is anonymized. The personal data leakage happens when a person can be identified from the combination of multiple anonymous open datasets".

58% of the respondents shows no concern about the possibility of data privacy breach. It seems in reality respondents have less concern on the possibility of data breach and the wording of choice situation exaggerate the chance. The policymaker should consider this fact in the interpretation of this study and further investigation is needed to conclusively determine the utility of data protection attribute.

Other than that, two attributes are considered insignificant by the citizens *"number of free citizen data skill training events"* and *"number of free engaging hackathon events".* However, the descriptive result shows that only 7% of the respondents have attended open education data events. It might have been difficult for respondents to assess their preferences for participation and engagement events (hackathon/data skills training) if they have never attended one.

Given the citizens reluctance to compromise the data protection attribute, government agencies have limited option for the implementation. Two recommendations are formulated to improve the existing open education data policy: 1) Collaborate with infomediary to provide services for citizens, and 2) Engage the citizens in a cost-efficient manner.

6.4. Limitation of the study

There are several limitations in the study:

Hypothetical situations instead of real situations

In the discrete choice experiments, the choice situations represent hypothetical situations rather than real situations. Therefore, it remains the question if respondents would make the same choices in real life situation.

Using secondary source for the policy context exploration

The policy context exploration is conducted through a desk research on the published policy documents of the government agencies responsible for open education data policy such as: Ministry of Education, Culture and Science (OCW), Education Executive Agency (DUO), and Education Inspection Agency (Inspectie van het Onderwijs). However, there is no primary source

in the form of direct communication with those respective agencies because the agencies do not accept the request for interview for student project.

Exclusion of cost attribute

The cost attribute is not included in the experiment due to the lack of information regarding the cost of implementation for each attribute from the secondary source. The information from the secondary source is highly aggregated and only shows the budget for the whole government agency. One of the respondents comments about the lack of cost attribute which will become one of the most important attributes for the government to compare between different alternatives.

Characteristics of respondents

The final survey is distributed among students who are currently attending Dutch higher education institution or recently graduated. The higher education students are targeted due to several reasons: 1) have relevant use case for the open education data which make them more likely to know about open data, 2) have relevant skills to use open education data, and 3) more likely to understand the term used in the survey with a proper explanation. The survey will gather different results if it is distributed in the general population, with more respondents who are not familiar with open data policy. In order to mitigate the homogenous characteristic of the respondents, the survey is distributed to the students with diverse study programs.

Limited selection of attributes

The attributes are selected based on three criteria: 1) Expected influence on an individual, 2) Societal relevance of the factor, and 3) Measurability in the discrete choice experiment. The attributes selected for the experiment are limited and may not reflect the whole possibility of attributes for citizens. For example, one of the respondents comments about using data skill training events as one of the attributes while there is another cheaper option such as creating online learning environment that can be freely accessed by the citizens.

6.5. Recommendations for future study

Using primary source information

In the limitation of the study, the exclusive use of secondary source in the study is discussed. Future study could contact the responsible government agencies and gain access for the primary source information. It is important to improve the realism of the survey and collect detail information that are not publicly available from the published policy documents.

Include cost attribute in the survey

If the future research able to gain access for primary information, it is important to include cost attribute in the survey. Each attribute implementation certainly comes with a price. Including the cost attribute will lead to a better decision for government to understand how citizen preferences will differ if they consider the cost to implement their choices. Combining it with different functions of government agencies that require the limited budget is also interesting. In this research, DUO do not have enough budget to implement the changes needed to comply with GDPR requirements unless it compromises the budget for the other functionalities.

Extend the research for different context of open data policy

In this research, the experiment is limited to open education data and targeted for higher education students. Future research can explore different policy context (e.g. open data policy in the government agency with less personal datasets such as Ministry of Agriculture or Ministry of Infrastructure and Water Management) or different respondents for open education data. For example, open education data policy for primary and secondary schools which targets the parents and pupils as the users.

Expand research with unobserved alternatives and attributes

In this research four attributes are used to generate the choice situations. However, in the reality there are many attributes that can be included or combined to make different alternatives. The portal-related attribute is omitted from this study because the limited implementation of open data portal in the Dutch open education data. However, if the future research explores the portal-related attributes of city open data portal, the attributes selected will be different from the attributes in this study. The attributes will focus on the functionality and features of the open data portal such as the visualization capability, collaboration and communication features, format of the data provided, etc. compared to the socio-technical perspective of this study.

Bibliography

Afful-Dadzie, E., & Afful-Dadzie, A. (2017). Open Government Data in Africa: A preference elicitation analysis of media practitioners. *Government Information Quarterly*, *34*(2), 244–255. https://doi.org/10.1016/j.giq.2017.02.005

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, *32*(4), 399–418. https://doi.org/10.1016/j.giq.2015.07.006

Charalabidis, Y., Alexopoulos, C., & Loukis, E. (2016). A taxonomy of open government data research areas and topics. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 41–63. https://doi.org/10.1080/10919392.2015.1124720

Chatfield, A. T., & Reddick, C. G. (2017). A longitudinal cross-sector analysis of open data portal service capability: The case of Australian local governments. *Government Information Quarterly*, *34*(2), 231–243. https://doi.org/10.1016/j.giq.2017.02.004

Dawes, S. S. (2010). Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly*, *27*(4), 377–383. https://doi.org/10.1016/j.giq.2010.07.001

Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, *33*(1), 15–27. https://doi.org/10.1016/j.giq.2016.01.003

Hanne Obbink. (2012, October 11). Er zijn wél slechte scholen | TROUW. Retrieved from https://www.trouw.nl/home/er-zijn-wel-slechte-scholen~a56a8c06/

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2), 132–157. https://doi.org/10.1086/259131

Lee, G., & Kwak, Y. H. (2012). An Open Government Maturity Model for social media-based public

engagement. *Government Information Quarterly*, *29*(4), 492–503. https://doi.org/10.1016/j.giq.2012.06.001

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly*, *32*(3), 323–332. https://doi.org/10.1016/j.giq.2015.05.006

Mangham, L. J., Hanson, K., & McPake, B. (2009). How to do (or not to do)...Designing a discrete choice experiment for application in a low-income country. *Health Policy and Planning*, *24*(2), 151–158. https://doi.org/10.1093/heapol/czn047

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics* (pp. 105–142). https://doi.org/10.1108/eb028592

Mouter, N., & Chorus, C. (2016). Value of time – A citizen perspective. *Transportation Research Part A: Policy and Practice*, *91*, 317–329. https://doi.org/10.1016/j.tra.2016.02.014

Mouter, N., van Cranenburgh, S., & van Wee, B. (2017a). An empirical assessment of Dutch citizens' preferences for spatial equality in the context of a national transport investment plan. *Journal of Transport Geography*, *60*, 217–230. https://doi.org/10.1016/j.jtrangeo.2017.03.011

Mouter, N., van Cranenburgh, S., & van Wee, B. (2017b). Do individuals have different preferences as consumer and citizen? The trade-off between travel time and safety. *Transportation Research Part A: Policy and Practice*, *106*(September 2016), 333–349. https://doi.org/10.1016/j.tra.2017.10.003

OCW. (2017). *Rijksbegroting 2018*. https://doi.org/ISSN 09217371

OCW. (2018). *Rijksjaarverslag 2017 VIII Onderwijs, Cultuur en Wetenschap*. https://doi.org/ISSN 09217371

openstate.eu. (2018). Amsterdam kicks off with a hackathon series about education – Open State Foundation. Retrieved April 22, 2018, from https://openstate.eu/en/2018/02/amsterdam-kicks-off-with-a-hackathon-series-about-education/

Petychakis, M., Vasileiou, O., Georgis, C., Mouzakitis, S., & Psarras, J. (2014). A state-of-the-art analysis of the current public data landscape from a functional, semantic and technical perspective. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(2), 34–47. https://doi.org/10.4067/S0718-18762014000200004

Reggi, L., & Ricci, C. A. (2011). Information strategies for open government in Europe: EU regions opening up the data on structural funds. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6846 LNCS, pp. 173–184). https://doi.org/10.1007/978-3-642-22878-0_15

Rijksoverheid. (2016). Data-Expedition report Ministry of Education, Culture and Science Rijksoverheid.nl. Retrieved April 22, 2018, from https://www.rijksoverheid.nl/ministeries/ministerie-van-onderwijs-cultuur-en-wetenschap/evenementen/onderwijsdata-onder-de-loep/data-expeditie

Safarov, I., Meijer, A., & Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity*, *22*(1), 1–24. https://doi.org/10.3233/IP-160012

Sayogo, D. S., Pardo, T. A., & Cook, M. (2014). A framework for benchmarking open government data efforts. *Proceedings of the Annual Hawaii International Conference on System Sciences*, (May 2010), 1896–1905. https://doi.org/10.1109/HICSS.2014.240

Susha, I., Zuiderwijk, A., Janssen, M., & Grönlund, Å. (2015). Benchmarks for Evaluating the Progress of Open Data Adoption: Usage, Limitations, and Lessons Learned. *Social Science Computer Review*, *33*(5), 613–630. https://doi.org/10.1177/0894439314560852

Thorsby, J., Stowers, G. N. L., Wolslegel, K., & Tumbuan, E. (2017). Understanding the content and features of open data portals in American cities. *Government Information Quarterly*, *34*(1), 53–61. https://doi.org/10.1016/j.giq.2016.07.001

Tim Berners-Lee. (n.d.). 5-star Open Data. Retrieved May 5, 2018, from http://5stardata.info/en/

Titah, J. H. R. (2017). Conceptualizing citizen participation in open data use at the city level. *Transforming Government: People, Process and Policy*, *11*(1), 99–118. https://doi.org/10.1108/TG-12-2015-0053

Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. *OECD Working Papers on Public Governance*, *NO.22*(22), 61. https://doi.org/10.1787/5k46bj4f03s7-en

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, *33*(2), 325–337. https://doi.org/10.1016/j.giq.2016.02.001

Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., & Dwivedi, Y. K. (2017). Open data and its usability: an empirical view from the Citizen's perspective. *Information Systems Frontiers*, *19*(2), 285–300. https://doi.org/10.1007/s10796-016-9679-1

Welle Donker, F., & van Loenen, B. (2017). How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, *10*(3), 284–306. https://doi.org/10.1080/17538947.2016.1224938

Zuiderwijk-van Eijk, A. M. G., & Janssen, M. F. W. H. A. (2015). Participation and Data Quality in Open Data use: Open Data Infrastructures Evaluated. *Proceedings of The15th European Conference on E-Government, Portsmouth, UK, 18-19 June 2015; Authors Version*. Retrieved from https://repository.tudelft.nl/islandora/object/uuid:c3e2530d-eaa2-409b-a700-b7107db7e159?collection=research

Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, *31*(1), 17–29. https://doi.org/10.1016/j.giq.2013.04.003

Zuiderwijk, A., Shinde, R., & Janssen, M. (2018). Investigating the attainment of open government data objectives: Is there a mismatch between objectives and results? *International Review of Administrative Sciences*. https://doi.org/10.1177/0020852317739115

Appendix

A. Summary of assessment studies

Study	Method	Object of Assessment	Measured aspects	Perspectives
(Afful-Dadzie & Afful-Dadzie, 2017)	Quantitative (survey)	Open data portal	Journalist preferences of data-related metrics: • data quality • data format • metadata • data availability • data integrity	Citizen (journalist)
(Chatfield & Reddick, 2017)	Quantitative	Open data portal	• open data provision • data format variety • open data policy intensity • entrepreneurial data services.	Government
(Thorsby et al., 2017)	Quantitative (scoring)	Open data portal	Open data portal features and content diversity. Features category: • content • help • policy • results	Government
(Welle Donker & van Loenen, 2017)	Qualitative	Holistic (data supply, data governance, user)	Data supply indicators: • Known • Attainable • Usable Data governance indicators: • Vision • Leadership • self-organizing ability • financing • open data stimulation • supply-user communication • G2G communication	Government and Citizen
(Lourenço, 2015)	Qualitative	Open data portal	Data disclosure characteristics: • quality • completeness • access and visibility • usability and comprehensibility • timeliness • value and usefulness • granularity • comparability	Government

(Zuiderwijk et al., 2018)	Quantitative (survey)	Relation of OGD initiatives and delivered benefits	Four categories of delivered benefits: • Operational • Technical • Economic • Societal	Citizen
(Safarov et al., 2017)	Qualitative (Literature Review)	Discussion of open data utilization in the academic community.	conditions for utilization: • quality of data • legislation/policy • skills • infrastructure • availability • privacy	Academic
(Zuiderwijk-van Eijk & Janssen, 2015)	Quasi-Experiments	Open data portal	participation mechanism and data quality indicators: • discussion messages • social media sharing • submissions of related items • wiki descriptions and discussions • data quality ratings • data quality reviews	Citizen
(Vetrò et al., 2016)	Quantitative	Open data portal	Data quality: • Completeness • Accuracy • Traceability • Currentness • Expiration • Compliance • Understandability	Government
(Weerakkody et al., 2017)	Quantitative (survey)	Citizen intention to use open data	• relative advantage • compatibility • observability • security risk	Citizen
(Ubaldi, 2013)	Qualitative	Holistic	• policies and law • technical • data governance • organizational • communication and interaction • political priorities • impact • data-related metric such as availability, quality, uptake, re-use.	Government
(Sayogo et al., 2014)	Quantitative	Open data portal	• data content • data manipulation capability	Government

			• participatory and engagement capability	
(Dawes et al., 2016)	Qualitative	Holistic	• policy and strategy • data publication and use • feedback and communication • benefit generation • advocacy and interaction among stakeholders	Government

B. Overview of choice situations

Design	Alternative 1				Alternative 2			
Choice situation	Mode of information presentation	Number of free engaging hackathon events	Number of free citizen data skill training events	risk of your personal education data exposed to the public	Mode of information presentation	Number of free engaging hackathon events	Number of free citizen data skill training* events	risk of your personal education data exposed to the public
1	in original form (as similar as possible to the source)	1 every 2-years	1 per year	1 incident per year	as static or interactive figures	1 per year	1 per year	1 incident every 3-months
2	as a service (e.g. an application such as studiekeuze123.nl or scholenopdekaart.nl)	1 per year	2 per year	1 incident per year	as static or interactive figures	2 per year	3 per year	1 incident per year
3	as static or interactive figures	2 per year	3 per year	1 incident per year	in original form (as similar as possible to the source)	1 per year	3 per year	1 incident per month
4	as static or interactive figures	1 per year	1 per year	1 incident every 3-months	as a service (e.g. an application such as studiekeuze123.nl or scholenopdekaart.nl)	1 every 2-years	3 per year	1 incident every 3-months
5	in original form (as similar as possible to the source)	2 per year	2 per year	1 incident every 3-months	as a service (e.g. an application such as studiekeuze123.nl or scholenopdekaart.nl)	1 per year	2 per year	1 incident per year
6	as a service (e.g. an application such as studiekeuze123.nl or scholenopdekaart.nl)	1 every 2-years	3 per year	1 incident every 3-months	as a service (e.g. an application such as studiekeuze123.nl or scholenopdekaart.nl)	2 per year	1 per year	1 incident per month

7	as a service (e.g. an application such as studiekeuze123.nl or scholenopdekaart.nl)	2 per year	1 per year	1 incident per month	as static or interactive figures	1 every 2-years	2 per year	1 incident per month
8	as static or interactive figures	1 every 2-years	2 per year	1 incident per month	in original form (as similar as possible to the source)	1 every 2-years	1 per year	1 incident per year
9	in original form (as similar as possible to the source)	1 per year	3 per year	1 incident per month	in original form (as similar as possible to the source)	2 per year	2 per year	1 incident every 3-months