The Dynamics of Open Science Adoption: A Choice Modelling-Based Approach

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in Engineering & Policy Analysis

Faculty of Technology, Policy and Management

by

Maarten de Graaf

Student number: 4236785

To be defended in public on August 10, 2020

Graduation committee

Second Supervisor & Chair	: Dr. Jan Kwakkel, Policy Analysis
First Supervisor	: Dr. Anneke Zuiderwijk, Information and Communication Technology
Advisor	: MSc Teodora Szep, Transport and Logistics

Summary

As open science has attracted attention within the academic world, the potential for innovative ways of information sharing offers multiple benefits, such as increased opportunities for collaboration an enhanced research transparency (Forrester, 2015). Often considered as the next step in scholarly dealings, the array on open science principles and their perceived benefits are omnipresent. Yet, their adoption by researchers themselves is lagging behind. The current body of research emphasizes on the conceptualization of open science, as well as inferring drivers and inhibitors from survey-based approaches. Nonetheless, insights into researchers' behaviour with regard to open science adoption are rather limited. With the importance of open science being acknowledged ubiquitously, the evaluation of policy levers to lead academics into this new era are rather unquantified. That is, although a vast amount of knowledge on the drivers behind open science is readily available; their relative importance to researchers remains largely unknown. The goal of this research is to unveil which factors catalyse (inhibit) the adoption of open science principles, as well as their relative importance. Besides a strong emphasis on exemplifying behaviour, this research seeks to quantify factors obtained through consulting descriptive literature. By doing so, a step towards understanding the relative importance of drivers and inhibitors may be taken.

Methodology

In order to achieve the preceding goals, this research attains a choice modelling-based approach. As prime means to elicit preference from the respondent group, stated-choice experiments were utilized as to present the opportunity to either employ open science principles or not to do so, under the variation of attribute levels. One may consider this research to contribute to exemplifying theory, as it seeks to quantify the current array of literature in terms of preference-based values.

Defining open science

The umbrella term that is open science harnesses a variety of domains (Fecher & Friesike, 2013). Although equally important to open science, the realm of open research occupies a central role in this research. We further subdivide open research into open data, open access and open source. As descriptive literature on open science adoption is omnipresent, a crisp, literature-based definition of the pillars that constitute those concepts was devised. Here, we dichotomize open data according to 1) social engagement 2) effort 3) recognition 4) data control and 5) data quality. Subsequently, we typify open science considering 1) social engagement 2) effort 3) visibility 4) recognition and 5) publishing costs. These factors represent categorically the main groups of drivers and inhibitors of open science adoption, based on contemporary literature.

Factors of influence

The adoption of open science principles by researchers depends on a variety of factors. A commonality between open data and open access is the negative relationship between adoption and perceived effort to openly share research data/publish through open access. Furthermore, increases in publishing costs and academic recognition were found to induce altered behaviour within the respondent group. Secondly, this study reports the existence of a variety of external factors that do drive (inhibit) behaviour, yet could not be included in the experiment itself. That is to say, certain barriers are in place that constrain the likelihood of researchers to engage in open science. In terms of open data, third party contracts and copyright concerns, as well as the competitive environment in which research groups operate, prevent them from openly sharing their research data. With regard to open access, subscription-based journals are considered more renowned and prestigious, rendering open access journals as a second choice. Furthermore, not every research discipline hosts a (qualitatively high) open access means of scholarly communication.

Policy implications

A multitude of policy implications can be inferred from the factors of influence. A strand of policy recommendations is directed towards unilateral action by research institutes. It is argued here academic bodies should attain a facilitating role in navigating the scientific community towards openness. Research institutions should 1) strive to minimalize the effort required to employ open means of publishing/sharing by providing adequate support, 2) recognize open-science-related doings is as important, as is mitigating the perceived level of effort and 3) devise financial frameworks, as to alleviate the shifted financial burden from customer-side subscriptions to production-side article publishing costs. Furthermore, we call for collaborative effort between stakeholders to ameliorate the external, inhibiting factors that are currently in place. As where research institutions do hold a crucial role in policy formation, they lack the ability to govern data contracts, improve open access journals, stipulate enhanced open data sharing agreements and impose grant requirements. National governments and the European Union hold greater jurisdiction and regulatory power and may facilitate large-scale open science principle adoption. By funding open science initiatives, hosting repositories and forming data policy as well as alliances with renowned publishers, they may incentivize open science principle on a higher level of aggregation. Grants and financers exhibit capabilities to stipulate open data and open access requirements within their agreements and are encouraged to do so. Although open access journals struggle with lower perceived quality, entering agreements with institutions, devising novel business models and reducing effort may propel them to the top of the scholarly communication realm.

A transition from traditional science to the realm of open science requires vast infrastructural changes, as well as a widespread adoption by researchers themselves. Here, an examination of the behavioral context induced unveiling a three-fold of main observations. First, prestige and career advantage exemplify researchers' behavior with regard to open science principle adoption. Both increased levels of recognition, as well as the unwillingness to publish in less renowned journals proved a strong indicator for the likelihood of scholars to tread in the direction of open science. Furthermore, ahead of open science being ubiquitously adopted, a variety of barriers is yet to be surmounted. Reshaping data agreements, lowering the bar for researchers to engage and mitigating the financial burden associated with open access publishing hold the potential to spearhead science into its new era.

Societal contributions & academic relevance

In terms of impactfulness, this research yields a set of novelties unprecedented by the current knowledge base. Departing from the realm of descriptive research, it has been sought to epitomize the behavioral dynamics of open science principle adoption. Along with a research method that is largely unseen within the current array of research, this adds to the literature a better of the trade-offs faced by academic actors in adopting open science. Furthermore, this study yields a quantifying layer to previously descriptive, qualitative factors. Zuiderwijk, Shinde & Yeng (Accepted) provide an overview of drivers and inhibitors of open data adoption. Here, we enhance those factors with an exploration of their relative importance, finding that effort required to publish open data is paramount towards adoption. Similarly, academic recognition and effort are highly significant factors, out of the five barriers stated by Björk (2004). On a societal level, this research is rendered to contribute towards a future in which science is common equity rather than an elitist dimension, inaccessible to the general public. By seeking to understand the behavior of those at the cradle of novel knowledge and advancement of society as a whole, this research is rendered to impact the approach governing (academic) bodies take with regard to open science and its adoption.

Contents

1. Introduction

With the rise of open science, new ways of conducting research have gained increased attention within the academic domain. A transition from traditional publications to open means of both publishing and open data repositories, culminated from the combination of research from different sources, could lead to an increased degree of collaboration within the academic world. Whyte & Prior (2011) denote this as an increase in the speed and efficiency of the research cycle. As where contemporary manners of publishing require a premium in order to be accessed, open science revolves around freely and openly sharing both research data and the resulting findings alike. Besides data sharing, the hypernym open science is not limited to open data sharing, but entails multiple facets (Fecher & Friesike, 2013). Therefore, it is crucial to appropriately scope the definition of open science opted for within this contribution. Here, we define open science as *the dissemination of scientific knowledge that is as wide as possible, free of charge to all users, and accessible online.*

The hypernym open science can be subdivided into three partitions; open education, open courseware and open research (de Jong, 2019). For clarity, a visualization of open science and its subordinate domains is available through Figure 1. As where the importance of each aspect is acknowledged, this research will emphasize on the latter, Open Research, marked green in Figure 1. This concept consists of three components, open access, open data and open source. Open data addresses research data sharing, which entails open sharing of data sets produced and utilized for research purposes (de Jong, 2019). Open access entails free availability of academic research to the general public by removing subscription-based barriers instilled in traditional publishing (Fecher & Frieske, 2013). Open source expands the realm of open data into applications. As where open data principles constitute research data itself, open source revolves around the manipulation of this data (Lindman & Nyman, 2014). This entails the open sharing of 1) source code developed to obtain experiment results and 2) the free availability of software and its underlying source code (Lindman, 2014).

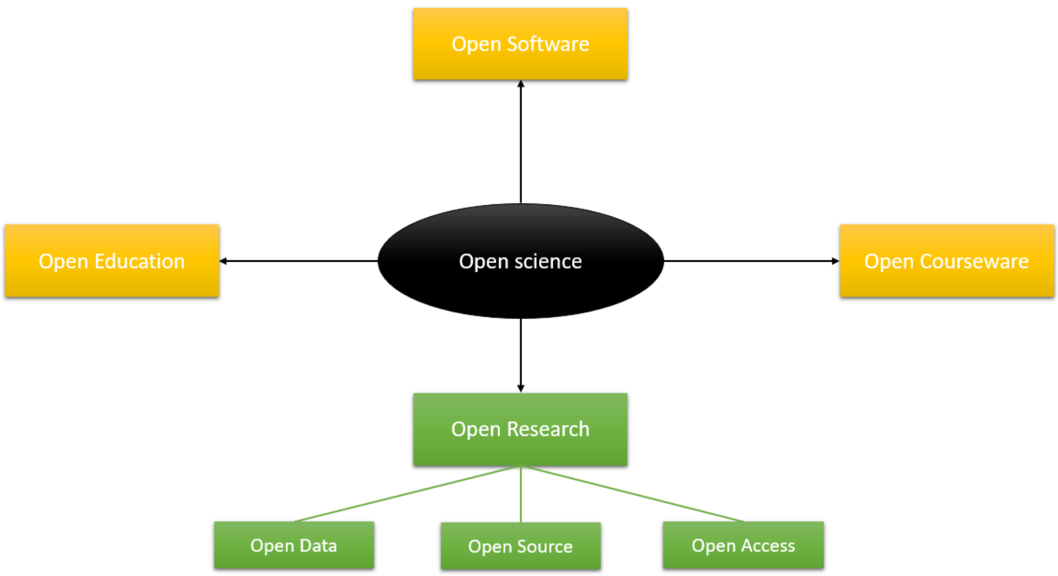


Figure 1.1 The dimensions of open science

Various, highly renowned academics have criticized the monetization of science and have held it accountable for the induction of various mishaps in behavioral conduct by academics (Sarewitz,

2016; Franzen, Rodder & Weingart, 2007). Sarewitz (2016) highlights that, with budgets allocated to scientific research skyrocketing up to forty-fold digits over the last fifty years, science has lost its accountability to society and created a realm of career-enhancing behavior for researchers, rather than experiments based on local and specific issues. A prime example of a monetization-induced mishap is Diederik Stapel, who has been charged with manipulating untruthful research data, seeing his contributions being declared fraudulent (Fraude hoogleraar Stapel 'verbijsterend', 2011). Due to the highly confidential type of research data, no corrective action occurred, rendering deceitful results to be assumed as credible ahead of the scandal's eventual surface. Hwang Woo-Suk, who claimed to have found breakthrough methods to cure cancer by means of stem cell treatment, further exemplifies misconduct in fraudulent data manipulation (Hwang and the Stem Cell Swindle, 2011). Eventually, his research data was found to be self-invented and thus fraudulent.
Such critical incidents have a powerful, detrimental impact on society trust in academic research (Franzen, Rodder & Weingart, 2007). Evidently, a lack of transparency may lead to catastrophic results and harnesses the risk of academic researchers doing as they wish to reach desired outcomes. Open science could play a crucial role in curbing risks associated to data opacity, with open data striving for collective sharing of research data and serve as a stage of peer reviewing one another's results (Kim & Adler, 2015). Besides transparency, open science enhances the potential for (cross-disciplinary) collaboration within the academic realm (Whyte & Prior, 2011). By speeding up the research cycle, both redundancy in terms of data collection through increased data availability and collaboration between previously unrelated disciplines and departments are readily facilitated (Whyte & Prior, 2011). On a career development note, researchers generally reap a larger degree of exposure from publishing through open access (Eysenbach, 2006). Horta & Santos (2016) list the number of citations as a predictor of academic career advancement. Hence, open science principle adoption does not only harness industry-broad advantages, but also offers personal incentives.

Both NGOs and governing bodies such as the European Union are considering the formation of open science policies and have already introduced advisory boards for related matters, with the Netherlands having their own National Program for Open Science (The Key Areas, n.d.). Besides creating awareness for open science, funding bodies are also employing policies to increase data quality (Open Science, n.d.). A linking study by Leonelli et al. (2016) explores the relationship between the U.K. Open Science landscape and research practice by academics. It was found that encouraging guidelines are in practice, but do not yet appropriately enforce the employment of open science principles by researchers, abstaining from prescribing how sharing might occur and is to be regulated (Kim & Adler, 2015; Fecher, Friesike & Hebing, 2011).
Landry, Traoe & Godin (1996) recognize a positive correlation between researchers' productivity and collaboration. Hence, the benefits of an integrated, cooperative process-enabling tool seem clear-cut. Nevertheless, researchers have found themselves only limitedly adopting open access publishing. In the Netherlands for example, it was found only 42% of research published in 2016 was made available in open access (Open Access Figures in the Netherlands, n.d.). Another insightful metric, the growth of articles available through open access, however generally positive, does not grow uniformly each year. This may indicate that various barriers currently obstruct further adoption by researchers. Amongst others, legal frameworks, IT-infrastructure, underlying business models and perceived rewards by researchers may form barriers for researchers to publish their research using open access (Björk, 2004). With regard to open data, limited adoption could be accounted for by the fact that researchers fear their data may be misused or misinterpreted, if not accompanied by proper explanation (Whyte, 2011; Zuiderwijk, Shinde & Yeng, forthcoming). In addition, open repositories harness the risk for researchers to see their data being published on, without them having been able to publish about the subject themselves (Whyte, 2011; Creaser, 2010; Forrester, 2015). From current

open data literature, it is not evident which types of researchers or datasets are susceptible to those risks. In addition, network externalities remain unassessed for individual brands of research. That is, certain disciplines might face more of the adverse effects, as where other disciplines would reap large benefits from open science implementations. How researchers perceive open science per discipline and which trade-offs they hold remains unknown (Forrester, 2015). From a policy perspective, ambiguity on the values and goals of stakeholders such as researchers remains highly undesirable. Not only does this complicate policy making in the sense of finding appropriate incentives to increase adoption, but more importantly, it obscures decision makers from understanding the underlying problems faced by researchers.

This research aims to **develop a behavioural understanding of researchers' adoption of open science principles (or lack thereof)**. With open research offering both personal drivers as well as drivers affecting academics as a whole, main emphasis lies on open data and open access. It is also sought to gain a further understanding of the relationship between open science, transparency and academic visibility. Although recognized as an aspect of open research, open source is not considered for experimentation, as its main relevance to open science stems from transparency enhancement (Lyon, 2016). Besides strong coherence, open data and open access more holistically cover open science as a whole, as where open source acts as a subsidiary aspect. The research objective can be considered as two-fold, 1) to investigate academic researchers' prioritization of benefits and barriers and 2) to investigate whether such prioritizations vary between disciplines. In terms of knowledge, it is expected this research will produce novel information on 1) which drivers and inhibitors are to be prioritized to enhance open science principle adoption by researchers, 2) whether this prioritization varies for different research disciplines and 3) add a quantitative dimension to current qualitative research. The latter serves to guide research institutions and governing bodies as to which factors should be focused on, as well as to provide the relative importance of factors included.

As main contribution, this research divulges the relative importance of attributes, specific either to a certain research discipline or in a general sense. Based on a preliminary literature search, it is expected attributions and returns from open science are influential factors for academics (Davis & Connolly, 2017). Furthermore, discipline-specific and cultural factors have also been indicated as potential predictors of open science adoption (Ali-Khan, Harris & Gold, 2017). As a result, those involved in decision-making within the field of academic open access may be informed on how to increase adoption by researchers. Both benefits and risks are assessed, as well as potential institutional agreements and incentive mechanisms that could be utilized to catalyse open data sharing and further reduce the risks involved. Besides the relative importance of factors, contributions consist of a set of policy recommendations, which may serve as decision-support for university board members concerned with open access, that need to create open data policies for their academic staff. It is noted actors within the system differ in terms of goals, objectives and values. Therefore, it is foreseen the contributions of this research are mere guidelines, rather than an exact solution.

2. Research Design

This chapter outlines the research design. It provides an overview of the research questions, which are addressed over subsequent chapters. First, the problem statement is defined according to knowledge gaps observed within the current array of literature. Secondly, the main research question and sub-research questions, distilled from the problem statement, are introduced. Subsequently, research outcomes and the choice of population is discussed as part of the research demarcation. To conclude this chapter, the research methodology is presented.

2.1. Problem Statement

Open access research is ubiquitous throughout academic literature and ‘open science’ has been labelled a buzzword by many authors. Nonetheless, a plethora of knowledge gaps appears eminent from an investigation of state-of-the-art literature. For this research, a triplet of those gaps have been elected as incipience to contribute to.

1. *A lack of research on the prioritization of factors for open science principle adoption*

As where an abundance of publications seeks to define the drivers and inhibitors of open science principle adoption, research on their relative importance is rather unseen (Eger, Scheufen & Meierrieks , 2016; Björk, 2013; Narayan & Luca, 2017; Leonelli, Spichtinger, & Prainsack, 2015). Consequently, factors of importance with regard to open science adoption are known; however, their reciprocal prioritization is highly uncertain. As current work predominantly adopts a global, descriptive approach, behavioural examination of open science principles by researchers is rarely seen within literature. Therefore, it remains significantly difficult to infer substantiated claims and develop sturdy frameworks that can be stipulated in academic policies. Furthermore, the absence of knowledge on the relative importance of factors obscures scholars from progressing towards fully understanding the underlying dynamics of open science principle adoption. That is, without determining which factors outrank others in terms of importance, research on novel factors of influence, as well as research on known factors that may potentially prove irrelevant, is void.

2. *Insufficient quantitative research*

The current array of research is predominantly of exploratory nature (Eger, T., Scheufen, M., & Meierrieks, D., 2016; Forrester, A., 2015; Leonelli, S., Spichtinger, D., & Prainsack, B.,2015; Bjork, 2013). Moreover, research outcomes report findings based on semi-structured qualitative research, but abstain from proposing a framework to stimulate open science adoption. In addition, the methods applied to open science-related research are primarily directed towards qualitative research. Structured, model-based approaches remain largely absent from literature. This leads to the inability to pinpoint behavioural explanations for the (lack of) open science principle adoption by researchers. Governing bodies are therefore unable to assess which factors should be preferred in terms of policy formation, since they lack quantified information on the impact of changes they might make to influence factors that are described qualitatively.

3. *Differences in adoption between disciplines are fuzzy*

Most research is dedicated to specific disciplines and if involving multiple disciplines, usually similar fields of study are considered (Schöpfel, Ferrant, André & Fabre, 2016; Whyte & Pryor, 2011). Hence, it is unknown how the adoption of open science may differ between disciplines and which drivers and inhibitors are of key importance to that particular academic research field. Although a niche of research has mentioned discipline-specific factors, greater efforts are required in order to validate their claims (Eger, Scheufen & Meierriks, 2016). Insights into disciplinary differences allows the field of open science research to develop itself in the direction of finding behavioural patterns specific to certain research disciplines. Rather than general trend observation, findings related to individual disciplines are more compelling with regard to understanding the dynamics of open science principle adoption. Furthermore, filling such knowledge gap enables decision makers to form clear-cut policy based on disciplinary differences, therefore enhancing its robustness.

2.2. Research Questions

Based on the literature review and the knowledge gaps in the field of study, the following research question is phrased:

"*Which factors explain the adoption or lack thereof of open science principles by researchers?*"

2.2.1. Sub-research questions

By means of the designated research approach in conjunction with the main research question, multiple sub-research question can be formulated:

1. *What factors drive and inhibit the adoption of open science principles by researchers?*

In order to assess the relative importance of factors, a preliminary step to distinguish the drivers and inhibitors behind open science adoption is required. Not only is this assumed to provide a clear overview of state-of-the-art knowledge on Open Science and its drivers, but also serves as an input for further experiment design. For answering this research question, a comprehensive literature search will be employed.

2. *What is the relative importance of factors that influence open science adoption?*

By distinguishing the drivers and inhibitors, no knowledge on their relative importance is gained yet. Therefore, a second research question is dedicated to finding the relationship between factors found relevant and their criticality. Research question 1 will serve as input for answering this research question, which will not only exclude obsolete questions from data gathering methods, but also provide a clear overview of the current state-of-the-art work in the field of quantitative assessment of Open Science incentivization. To stipulate a metric for relative importance, stated choice experiments will be applied as a means of data gathering to be later integrated into a comparative analysis of factors.

3. *How does the relative importance of attributes vary across research disciplines?*

A secondary interest with regard to Open Science adoption are the differences between research disciplines. Universities are often partitioned into multiple faculties, with the number of different departments being increasingly sizable for large institutions. For academic decision makers and university boards of those institutes, it is of key importance to incorporate field-specific regulations into their policy to enhance adoption. A cross-reference of stated choice experiments, performed on subjects from different research disciplines, distinctive patterns can be explored and further evaluated.

4. *What are the policy implications of the attributes' relative importance?*

Findings gained during stated choice experiments, the method for comparative data gathering, are to be translated into a ranking system. The penultimate goal of this research is to develop a typology for a quantitative assessment for drivers and inhibitors for Open Science adoption. Experiment results yield a vast amount of knowledge on researcher preferences, which then can be used to derive relative rankings of specific factors. In order to do so, data analysis methods will be applied to the experiment data set. This research question represents the process of data analysis and corresponding conclusions

2.2.2. Sub-question matrix

To provide an overview of sub-research questions and their associated methods, a sub-question matrix is available through Table 1.

Sub-research question	Method
RQ1: What factors drive and inhibit the adoption of open science principles by researchers?	Literature review (syetematic)
RQ2: What is the relevant importance of factors that influence open science adoption?	Stated choice experiments
RQ3: How does the relative importance of attributes vary across research disciplines?	
RQ4: What are the policy implications of the attributes' relative importance?	Data analysis

2.3. Methodology

In order to answer the research question and corresponding sub questions, various methods, tools and datasets are required. This section contains an outline of those required methods, along with

2.3.1. Research approach

Since the aim is to gather comparative data, employment of a non-descriptive research approach is required. Exploratory research in the field of open science is ubiquitous throughout literature (Fecher & Friesike, 2013; Whyte & Prior, 2011; Mosconi et al., 2019). Therefore, the novelty of insights gained by employing a qualitative approach is limited). Since a prime gap within current research lies within an individualistic, behavioural dichotomy for risks and benefits experienced by researchers in open science adoption, the research approach should facilitate the extraction of such information. The main research question seeks for the conceptualization of the relative importance of factors influencing open science adoption. Methodologically, the research design should incorporate means capable of producing such results.

Stated choice experiments are a viable method to infer values from preference-based research. Rather than survey-based approaches, which are often based on dissimilar, non-comparative questions, stated choice experiments elicit preferences by the subject under study. The relative importance of attributes is inferred from presenting the respondent with a set of choices with variance across different sets (Klojgaard, Beck, Sogaard, 2012). Given the fact individualistic behavioural patterns are to be examined, stated choice experiments suit the purpose of eliciting researchers' preferences. Haider (2002) states choice models enforce respondents to think in a trade-off manner and enhance their capabilities of expressing preferences in a relative sense. Since the adoption of open science principles is inevitably based on trade-offs, one may choose to employ open data sharing or open access publishing and reap its associated benefits, this suits the goals of this study. Furthermore, choice-based questions hold enhanced potential for the collection of behavioural data over revealed preference experiments (Adamovicz & Louviere, 1998).

It is noted that a vast amount of desk research is required for experiment design. That is, choice set design are to be based on existing literature. Since the effectiveness of stated choice experiments are highly dependent on experiment design, research will predominantly focus on securing suitable choice alternatives. Moreover, the research approach is multifaceted by nature and involves both experiments, theory-based research and framework stipulation. It is expected to yield results in both a descriptive and bifurcating manner.

2.3.2. Data

The main research question is directed towards attaining a model-based approach towards open science adoption by researchers. For the research question attempts to address pattern recognition to discover preferences amongst researchers, a vast amount of data is required. Contemporary research on open science adoption does not cover neither comparative nor quantitative analysis. That is, no data on the relative importance of influential factors towards open science adoption is available as of now. Therefore, *comparative data on relevant dimensions for open science* is the first requirement to answer the research question in terms of data. It is noted that, in order to obtain such data and to conduct experiments, it is first necessary to distinguish which factors are sufficiently significant to include for analysis. According to this information, experiments can be designed and research questions can be answered. Furthermore, sub research question three, directed at discipline-based differences, stipulates a multidisciplinary approach. Therefore, data collection should yield a heterogeneous dataset in terms of participant faculty.

2.3.2. Research Demarcation

This section further demarcates this research in terms of population, space and disciplines. Moreover, the population participating in experiments is further described and characterized.

Choice of population

Since main emphasis lies on the open research aspect of open Science, a crisp specification of the desired research population is strictly necessary. With the criticality of open science principle adoption by academics, it is evident they should be featured as main group of interest. A general, open-to-everyone population would not suffice and more importantly, inject noisy data into the set of responses gained. Defined as broadly as possible, the target audience for conducting data gathering is therefore denoted as academics.

It is noted this initial group is rather large and disregards cultural as well as disciplinary differences between academics. Therefore, it is rendered paramount to distinguish further the desired set of respondents.

Spatial boundaries

In order to obtain perceptive experiment results, it is necessary to establish proper spatial boundaries. It is widely acknowledged cultural differences lead to different approaches to organizational culture. As Hofstede (2010) describes, schools and academies are affected by those differences as well. That is, it is to be expected that results gained from this research may differ from data gained if conducted elsewhere. Schöpfel et al. (2016) further highlights those differences by finding that barriers found by Björk (2004), which is considered a pioneering study in the field of open science principle adoption research, were largely absent at French universities.
In order to impede such discrepancies within research data, *academics participating in this study must be affiliated with Dutch-governed institutes* is imposed as a limitation, since they are assumed to belong to a similar cultural group and thus exhibit complementary organizational culture. Besides complementary organizational culture, limiting the geographical boundary to the Netherlands will

ensure participants are members of institutions subject to similar privacy laws and legal frameworks, which are proven influential during literature review.

Disciplinary boundaries

A prevalent knowledge gap is the lack of research on differences in adoption between research disciplines. Therefore, a comparative analysis of how drivers and inhibitors vary per sector would be a valuable contribution to the existing array of literature. Therefore, data collection involves specific targeting of disparate research disciplines. Due to the time available for this research, it was chosen to aggregate fields of study into their core branches, as specified by McGinn (2012). Those include 1) biology 2) technology 3) sociology and 4) economics.

2.3.3. Research Methods

To collect the required data, multiple means are viable for application. First, a thorough literature review unveils aspects relevant to open science adoption by researchers, as found through de facto publications. Furthermore, contemporary literature serves as a basis for experiment design and supplies qualitative elements to include in quantitative analysis as pursued by this research. Literature search is assumed to serve as a means to demarcate the system and form a basis for experiment design.

The main research method applied to answer the research questions are stated choice experiments. Predominantly, this research is directed towards distinguishing the relative importance of factors influential towards open science principle adoption. Therefore, non-comparative data collection methods such as surveys do not suffice as no ranking can be distinguished from responses. Stated choice experiments are designed to elicit preferences from participants when presented a set of attributes with certain levels assigned to them (Klojgaard, Bech & Sogaard, 2012). By means of translating adoption factors to attributes and creating levels to infer the likelihood of respondents to adopt open science principles, the comparative data required to achieve research goals can adequately be obtained.

The bipartite composition of research methods is expected to yield both inputs for experiment design, as well as a method to convert this design into comparative data. The latter can further be exploited to answer multiple research questions and may eventually be manipulated into a deterministic framework.

2.3.4. Tools

Multiple tools are necessary in order to gather, process and manipulate the data as well as answer both the main research question and relevant sub questions. Academic repositories allow for finding relevant publications on open science adoption and hence facilitate a thorough literature review. As TU Delft, the institution through which this research is executed, provides access to such means, it is foreseen this tool is readily available for consultation during each part of the research. This repository is an instanced version of WorldCat and integrates the digitalized content of TU Delft's library, its subscriptions and scholarly communications published within its ranks.

For the design and execution of stated-choice experiments, various additional tools are required as well. First, experiments are performed within an experimental space. That is, the conceptual experiment design should be stored either physically or digitally as to publish it to participants. In addition, experimentation requires an environment in which participants are able to engage in the experiment and have their data stored for further evaluation. Participants interact with the experiment to Qualtrics, an online survey development tool. Stated-choice experiment design was made of the Ngene software package. Multiple means were provided to store data results; main use was made of an encrypted project drive on a remote TU Delft server (Surfdrive). It is noted both software packages are subject to a license fee. The licenses required for experiment design and execution were provided to the author by Delft University of Technology.

In order to distinguish the relative importance of factors, multiple data analysis operations need to be applied. In order to do so, various tools are to be utilized as to examine the dataset. Considering the skillset and prior experience of the author of this research, RStudio has been chosen as means to train, examine and manipulate the datasets. For data analysis, various RStudio libraries were utilized. Psych served as a means to obtain (descriptive) statistics, as where plyr was employed to perform data frame manipulation as to clean and prepare the dataset for further examination. Furthermore, Apollo, a choice modelling RStudio package, was utilized for analysis of stated choice data (Hess & Palmer, 2019). For visualization purposes, datasets were transferred to PowerBI and incorporated into numerous graphs and figures.

2.4. Summary

The research design provides an overview of the main inputs for investigating open science adoption by researchers. Three knowledge gaps were observed, namely; 1) the lack of research on the prioritization of factors for open science principle adoption 2) insufficient quantitative research and 3) a fuzzy understanding of the differences in adoption between research disciplines. Four research questions address these knowledge gaps, attempting to unveil novelties as to close the gap. In order to do so, this research applies two mains methods, which cascade into each other:

1. Literature review: yields an overview of drivers and inhibitors to be later used during experimentation
2. Stated-choice experiments: holds the relative importance of factors found throughout literature, as well as differences in adoption between disciplines.

The research demarcation restricts participation to members of or affiliates with Dutch universities, as to rule out cultural differences. Furthermore, heterogeneity is pursued within the respondent group in order to gain insights into disciplinary differences. A plethora of tools is utilized to arrive at eventual policy recommendations, but predominantly include data analysis and visualizations solutions (Rstudio, Rstudio packages & PowerBI).

3. Research Background

3.1. Introduction

The construction of a research background is an instrumental step towards generating and executing the research's experimental phase. This chapter contains an overview of the various definitions of open science, as well as its benefits and a first exploration of barriers to open research. Such definitions serve to provide an initial framework for further experimentation and they form the pillar, which this research is built upon. Subsequently, the drivers and inhibitors of open data, open access and open source are scrutinized, as to answer **research question 1:**

What factors drive and inhibit the adoption of open science principles by researchers?

As a closing remark, this chapter frames the observed drivers and inhibitors according to the main research objective, by means of the research scope. The research background can be considered as framework for experiment design and serves as impetus to (stated-choice) experimentation.

3.2. Literature Review

Vast amounts of literature have been published in the field of open science. The literature review section is structured as follows: 1) a review of open science principles and different paradigms, 2) the value of open science and 3) barriers to open research.

3.2.1. Literature review approach

A methodological approach to reviewing literature is critical to success for any academic research (Webster & Watson, 2002, pp. 48-49). Not only does an examination of the current array of literature provide a solid theoretical foundation for the proposed study, it also helps the researcher understand the existing body of knowledge and identify the necessity for new research (Levy & Ellis, 2006). Due to the sizable collection of research on open science, it is deemed critical to clearly structure the literature review and predefine both the process and desired outcomes beforehand. Levy & Ellis (2006) propose a sequential, "input-processing-output"-based approach, which denotes 1) inputs 2) processing and 3) output as disparate steps within a literature search. As an input, the following goals for the literature review were distinguished; 1) to define open science as key concept, 2) to identify the benefits of open science versus traditional approaches and 3) to distinguish the barriers to open research adoption.
As a secondary input, the repositories consulted for the literature review are ought to be made explicit. The following databases have been examined to obtain relevant papers: Semantic Scholar, Google Scholar and the TU Delft library database. The latter is an integrative WorldCat implementation that harbours publications by TU Delft and all of its journal subscriptions.

In terms of processing, only one criterion was posed to literature search results. Levy & Ellis (2006) propose to attain an approach that goes beyond assuming peer-reviewed articles as a benchmark for quality assurance. That is, not merely a peer-reviewed status suffices, but the status of the underlying journals and conference should serve a strong indicator as well. However, it was decided here to relax this constrain and consider peer-reviewed articles as appropriate for inclusion, may the piece of research be considered as a fruitful contribution towards one of the specified literature review goals. Furthermore, conference proceedings are also rendered as a credible source of information and they were therefore consulted as well, if considered to contain valuable knowledge.

Operational search methods used during the literature review include keyword search, backward search and forward search. A plethora of keywords was applied to conjure the literature review; an overview is available through Appendix IV. It is noted that a part of the literature was consulted throughout an early stage of this research and therefore deviated from the structure described above. Nonetheless, the input of keywords and backward search were applied in order to compose an initial version of the literature search.

The main goal of the literature review is to **create an overview of the current body of knowledge, define open science according to the main research goals and explore the value of open science, as well as general barriers to adoption**. In addition, the literature review serves to **distinguish the drivers and inhibitors of open access, open data and open source respectively**.

3.2.2. Defining open science

A formal definition of Open Science remains subject to a degree of ambiguity. Due to its multilateral nature, formulating an all-embracing, crisp description remains a challenge in itself. Research efforts towards conceiving an encompassing definition are prevalent throughout literature. Murray-Rust (2008) conducted a pioneering study on open access in science, exemplified through a case study at the American Chemistry Society. The definition of open science differs between literary sources and evokes different thoughts. Murray-Rust (2008) describes open science as research being freely available on the public Internet, to be used as desired. Schöpfel et al. (2016) words open science in a democratizing way, stating that open science is a transforming process to enhance transparency, collaborative efforts between researchers and bridge the gap between science and society. Fecher & Friesike (2013) distinguish three more definitions of open science, which can be found in Figure 3.1. They add a measurement school to the realm of open science, calling for an alternative metric for scientific impact. Current metrics are not tailored towards the innovation brought by open science and therefore need reworking (Fecher & Friesike, 2013). The pragmatic school presented by Fecher & Friesike (2013) revolves around capitalizing on efficient means of collaboration and enhancing knowledge creation. That is, open science should be seen as an enabler for more integrated research efforts and infuse synergy into currently separated research institutes. For this research, the definition of open science predominantly focuses on pragmatic and measurement factors. Although the importance of democratization and public availability is acknowledged, it is argued that results that are more auspicious stem from seeking incentives in the field of different measuring systems and process optimization. Nevertheless, other definitions of open science are thereby not disregarded and if found influential, factors from other domains will be considered.

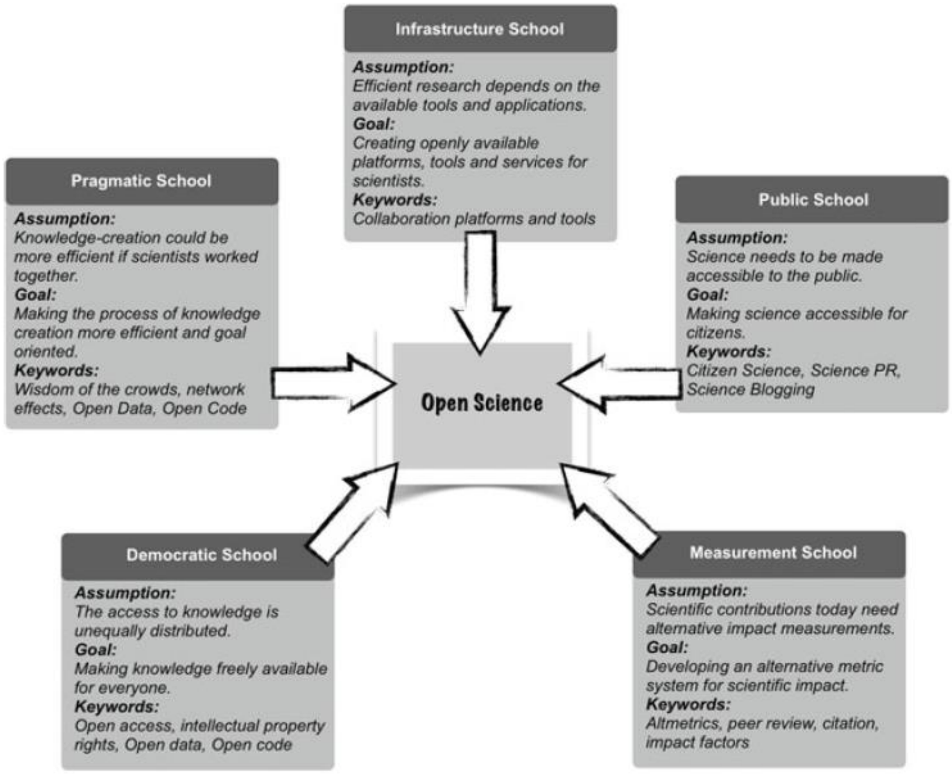


Figure 2.1 Five schools of open science by Fecher & Friesike (2013)

3.2.3. Open science value

Although variance exists on what open science entails, its value has been widely acknowledge throughout literature (Murray-Rust, 2008; Mosconi et al., 2019). A pioneering study by David et al. (2009), juxtaposes the advantages of open science according to a multilateral approach. It states that, from a functional perspective, open science holds incentive capabilities as to accelerate validation of findings, reducing excess duplication of research efforts, whilst stressing the importance of social organization to reap the benefits of open science. Whyte & Pryor (2011) further stress validation advantages with research data being openly shared, stating that the future of research should be guided by open data as the new common denominator within the scientific community. As where the importance of open access is recognized as critical towards the future of collaborative research, awareness tends to be less spread than necessary for adoption (Murray-Rust, 2008). Survey-based studies confirm this lack of awareness. Predominantly, researchers are unaware of the potential of open access systems and hence do not use them (Schöpfel et al., 2016). Moreover, cultural and disciplinary differences affect open access adoption as well. That is, certain countries differ in terms of publishing culture and are hence more likely to adopt open access repositories (Eger, Scheufen & Meierriks, 2016). In addition, open access adoption has been found to be highly discipline dependent as data sharing is more common in certain fields than in others.

3.2.4. Barriers to open research

Multiple studies have recognized the presence of barriers to entry for researchers to adopt open science networks. Björk (2004) identifies several barriers to change that hamper change within the system. It is argued a variety of paradigms exist within open access itself: 1) open access journals, 2) subject-specific repositories and 3) institutional repositories. Those breeds of open access typologies are of disparate nature by design. As where the former considers an open access approach to traditional journals, the second concerns repositories dedicated to one specific discipline and the latter encapsulates repositories developed for an institution as a whole.

Research conducted by Björk (2004) serves as a basis for further research within the field of study. Case studies regarding open access adoption unveil potential impediments and they have confirmed findings stated earlier (Whyte, 2011; Creaser, 2010; Forrester, 2015). However, various case studies found misconceptions with regard to open access as well. That is, misconceptions on open access usage and potential were found repeatedly by conducting semi-structured qualitative research (Narayan & Luca, 2017).

With substantial research dedicated towards investigating barriers to open access adoption, efforts towards identifying incentives are prevalent as well. A strong emphasis on stakeholder-driven open science can be recognized throughout literature. Stakeholders – or researchers, for this instance – exhibit various concerns with regard to open science. Mitigation of those preoccupations is paramount in order to reach higher amounts of researcher participation (Ali-Khan, Harris & Gold, 2017; Leonelli, Spichtinger & Prainsack, 2015). Ali-Khan, Harris & Gold (2017) conducted research on open science in biomedical research and found various additional constraints, introduced by the highly competitive, discipline-specific nature of this field of study. As where intellectual property is not of mere importance, proper attribution and returns from open science do incentivize researchers to apply such principles. Other studies acknowledge this view, exhibiting similar results (Leonelli, Spichtinger & Prainsack, 2015; Davis & Connolly, 2017). Schöpfel et al. (2016) restrains incentives proposed by other work within the field and presents a contradictory study that found barriers described by Björk (2004) to be largely absent respecting French universities. That is to say, ambiguity with regard to the underlying dynamics of open science adoption exists.

3.3. Drivers & inhibitors of open data adoption

A vast array of literature has been directed towards distinguishing the drivers and inhibitors of the adoption of open data principles. The current body of knowledge on influential factors can be further subdivided into different themes to compose a general overview. Over the course of the following sections, one may notice that factors do not exclusively belong to one specific aspect. For instance, social influence is relevant for both open data and open access.

Within the field of open data, Zuiderwijk, Shinde & Jeng (forthcoming) note several, principal themes. Increasingly relevant for experimental design were found:

1. Policies and voluntariness: whether open data policies are in place and whether sharing is considered voluntary;
2. Facilitating conditions: anything that can facilitate open research data sharing and use such as academic policy, IT and legal support;
3. Trust: the level of confidence in open data sharing and use;
4. Social influence and affiliation: factors related to social influence and affiliation that influence whether a researcher is driven to share and use open research data;
5. Effort: the effort needed for a researcher to openly share or use research data,
6. Legislation and regulation: the influence of factors related to legislation and regulation.

The themes listed above do not imply the exclusion of other factors, as the importance of demographic factors is recognized as well. The following sections will dissect the factors listed and how they may either drive or inhibit the adoption of open data principles.

3.3.1. Policies and voluntariness

Policies that either encourage or enforce researchers to openly share data play a sizable role in increasing the adoption of open science. The influence of both governing bodies and funding policies has been acknowledged throughout literature (Kim & Adler, 2015; Fecher, Friesike & Hebing, 2011). However, it is noted policies and standards are often inconsistent across disciplines and institutions, with data sharing often not being common practice (Kim & Adler, 2015). Nonetheless, Piwowar & Chapman (2008) describe a positive correlation between the strength of journals' data sharing policies and the degree to which scientists publicly deposit data into repositories. Wallis, Rolando & Borgman (2013) add to this statement by establishing that funding agency expectancies in terms of data sharing are a significant factor with regard to publishing datasets online. It can therefore be concluded policy may form a strong driver behind open data sharing by researchers, although it is currently not being ubiquitously applied effectively by governing bodies.

Zuiderwijk, Shinde & Jeng (forthcoming) acknowledge the impact of peer pressure on the likelihood to distribute data openly. That is, if a researcher experiences an environment in which open data is considered as a norm by others, it is more common for the researcher to adopt open data principles himself. Contrarily, Ceci (1988) raises concerns with funding under wide adoption of open data. That is, with datasets being widespread throughout the academic world, researchers may find themselves struggling to obtain funds for their study or see their funds being sizably reduced, due to the fact data gathering is rendered obsolete for a particular study. Academics are less likely to adopt principles that will possibly deteriorate their funding opportunities (Mooney & Newton, 2012).

3.3.2. Facilitating conditions

The creation of a facilitating environment finds itself at the core of either inhibiting or driving researchers to share their data (Zuiderwijk, Shinde & Jeng, forthcoming). A crucial facilitating condition for open data sharing is the availability of proper infrastructure. Wallis, Rolando & Borgman (2013) state that, especially outside of the big, science fields with thoroughly developed community standards, infrastructure is often lagging behind. The lack of infrastructure hampers openly sharing research data (Wallis, Rolando & Borgman, 2013). With the introduction of open data infrastructure comes a growing need for sustainability and flexibility of those repositories. Researchers are less likely to adopt a new means of sharing if unsure whether this enables long-term access (Zuiderwijk, Shine & Jeng, forthcoming).

Arzberger et al. (2006) further underlines the importance of technological infrastructure to support open data sharing. In addition, it notes that this infrastructure should be robust to serve for a long-term period. However, the technological requirements are rendered most surmountable in comparison to factors from other domains. It is noted technical aspects can both drive and inhibit open data sharing. That is, the availability of proper tools may encourage researchers to condone sharing, as where a lack thereof may cause researchers to abstain from employing open data principles (Arza & Fressoli, 2017). Arzberger et al. (2006) scrutinizes financial planning in research, as it notes misallocations of budget between actual research and data management, which undervalues the need for clear-cut data infrastructure governance.

3.3.3. Trust

Trust can be an influential driver and inhibitor for open research data sharing. The dimension of trust is multifaceted and does not only concern researchers' trust in the open data portal itself. Moreover, it concerns the trust of peers and society in research findings and the trust of open data users' in individual researchers themselves (Zuiderwijk, Shinde & Jeng, forthcoming).

Ample research suggests that various trust-related benefits may drive open data sharing by researchers. Kim & Adler (2015) mention enhanced data credibility can be achieved from data sharing, as well as increasing the likelihood for existing studies to serve as input for novel ones. Furthermore, trust in knowledge sharing positively influences individuals' knowledge sharing behaviour. That is, one is more likely to share data if one has a higher trust in means to share data openly (Kim & Adler, 2015). Fecher, Friesike & Hebing (2015) add to this by stating that an enhanced understanding of how users may utilize openly shared data, researchers' are more inclined to open up their data. Clear regulations on data ownership with researchers being able to influence how their data is being used, is listed as a driver for adoption as well (Fecher, Friesike & Hebing, 2015). Moreover, open data inhibits concerns related to data falsification and data fabrication (Tenopir et al., 2011). This may not be a researcher-side trust issue, but is considered to belong to this domain nonetheless.

Various trust-related inhibitors hamper open data sharing by researchers. One of the main concerns is the fear of misinterpretation and misuse of data after publishing (Fecher, Friesike & Hebing, 2015; Tenopir et al., 2011; Enke et al., 2012). That is, academics fear their data may be usurped for faulty purposes if not accompanied by proper explanation. Raw data itself does not contain experiment conditions under which it was gained, as well as the purpose and methods that were employed. Someone may therefore misinterpret data (Fecher, Friesike & Hebling, 2015). Subsequently, loss of control over unpublished data in publicly accessible online databases forms a strong inhibitor as well (Enke et al., 2012; Tenopir et al.; 2011; Borgman, 2012). With concerns over data ownership after sharing through public databases, researchers find themselves worried about how they retain control over their first-hand findings. Predominantly, researchers stipulate retaining first rights to publish results as main condition for data sharing (Jillian et al., 2013). This shows the concern they have with data ownership.
With regard to open data usage, difficulties in establishing data credibility inhibit employing second-hand data for research purposes (Jillian et al., 2013). This inhibitor is a double-edged concern, since not only will open data usage be encumbered by a lack of trust, but individual sharing as well. That is, if one does not trust data obtained through public repositories, one will perceive a low credibility score will be assigned to one's own research data if it were shared through open data means. Fecher, Friesike & Hebling (2015) adds that knowledge on who may potentially use openly published data leads to enhanced trust and a higher likelihood to share openly. Jillian et al. (2013) strengthens this claim by stating that researchers are willing to make a substantially larger effort to provide data to researchers they have worked with closely.

3.3.4. Social influence and affiliation

Multiple studies stress the importance of social influence and affiliation with regard to open research data. Drivers observed in the social domain include social responsiveness, peer pressure and attitudes with regard to data sharing (Zuiderwijk, Shinde & Jeng, forthcoming). Kim & Adler (2015) describe normative pressure as an important factor supporting data sharing within communities. If norms within a research community are to share data, individual academics are likely to abide by those standards and share data as well. Another driver for open data sharing is peer pressure, causing researchers surrounded by open data advocates to engage in public publishing (Piwowar & Chapman, 2015). Subsequently, normative pressure can also relate to pressure by journals or funding bodies (Kim & Adler, 2015). Despite this also being considered a policy-related factor, it is considered to be related to attitude nevertheless.

The social inhibitors for open data sharing observed throughout literature are limited to one. Sayogo & Pardo (2013) mention the culture of open data sharing as limiting to adoption. That is, if academic

recognition, promotion and glorification stems from publishing rather than open data sharing, researchers are more inclined to publish rather than to share. Subsequently, if organizational culture prohibits data sharing due to i.e. security restrictions, open data may not be able to penetrate deeply into that institution. Although intersecting the policy domain, this argument mainly revolves around the underlying organizational culture.
It is noted social drivers may also apply inversely. If either normative pressure or peer pressure is biased towards not sharing, researchers may find themselves encouraged to not share data openly (Zuiderwijk, Shinde & Jeng, forthcoming).

3.3.5. Effort

The likelihood of adopting open data principles is related to the perceived effort for researchers to share their data, as well as the effort required to obtain openly shared data (Zuiderwijk, Shinde & Jeng, forthcoming). Several literary sources mention perceived effort negatively correlates to open data sharing adoption and is therefore a powerful inhibitor (Harper & Kim, 2018; Kim & Adler, 2015). Stated simply, if something is perceived as being tardy and time-consuming, one is not likely to adopt it. Subsequently, data may exist, but unable to be found across an abundance of data repositories (Campbell, 2015). Moreover, the amount of repositories may strike researchers as overwhelming, rendering them intimidated and confused as to where to obtain data they are in search of (Campbell, 2015).

In the sense of effort, researchers are driven to employ open data sharing by savings with regard to reproduction and research cost reduction. Campbell (2015) states that open data avoids duplication of effort. If datasets are available online freely, novel applications of that data may be found without re-collecting them. This cascades into lower research costs, since a major part of data collection becomes obsolete under open sharing of required data sets (Campbell, 2015). Furthermore, researchers are more likely to engage in open data sharing when they perceive the likelihood their data will be re-used as high (Curty et al, 2017). If organizational culture supports data sharing with appropriate (data) management and tools that meet the requirements of researchers, required efforts are gradually reduced, which would drive adoption (Sayogo & Pardo, 2013).

3.3.6. Legislation and regulation

Legislative frameworks and regulatory measures may serve as a driver or inhibitor towards open data sharing (Zuiderwijk, Shinde & Jeng, forthcoming). A subset of drivers and inhibitors belongs to the realm of policy formation. Tenopir et al. (2011) states that the vast amount of data challenges faced in open data must be addressed by comprehensive, transparent data policies. Huang et al. (2012) enhances this by stressing data-related policies must be expanded as widespread as the organizational policy level. Fecher, Friesike & Hebing (2015) emphasize that journal policy could also be employed to either drive or inhibit open data sharing. That is, if policies from the groups mentioned above either encourage or enforce open data, it will drive adoption. Inversely, it will inhibit researchers from openly sharing their research data.
Legal frameworks may impose openly sharing research data. This may include copyrights or licensing issues, implying researchers are forced to choose between an abundance of platforms with different datasets available, eventually confusion the user (Fecher, Friesike & Hebing, 2015). In addition, ownership and confidentiality may further inhibit open data sharing (Fecher, Friesike & Hebing, 2015; Kim & Adler, 2015).

Privacy concerns also belong to the realm of legislation and regulation. If personal data is involved in open sharing, privacy concerns may be raised (Harper & Kim, 2018; Kim & Adler, 2015). Several data-protection laws are in place to prohibit the sharing of privacy-sensitive data, such as the General Data Protection Regulation employed by the European Union. Although data may be anonymized as to forego identities, re-identification tools to reverse this process are widely available (Henriksen-Bulmer & Jeary, 2016). Privacy may therefore be a strong inhibitor to opting for open data, despite the existence of measures to mitigate those concerns.

3.4. Drivers & inhibitors in open access

As where open data solely concerns the realm of data sharing and free availability of datasets, open access surpasses this realm to incorporate free publishing of journal articles. It is noted drivers and inhibitors from the domain of open data correspondingly apply to open access as well, which stems from uniformly shared open science principles.
In order to prolong the classification-based approach applied throughout past chapters, five dimensions are selected to categorize drivers and inhibitors for Open Access adoption. Björk's (2004) influential study served as a baseline in the development of the framework used here. The following dimensions are recognized within open access:

1. Social influence and affiliation: factors related to social influence and affiliation that influence whether a researcher is driven to publish through open access means.
2. Facilitating conditions: the availability of digital tools i.e. platforms and repositories to publish articles openly. This also includes overall support available to researchers to adopt open access and perceived effort to publish using open access.
3. Business models: concerns the development of monetization of open access publications.
4. Recognition: the degree to which academics may gain from publishing in either open access journals or open repositories.
5. Visibility: the extent to which researchers are able to gain exposure by publishing through open access.

It is rendered these dimensions adequately cover and categorize drivers and inhibitors faced by researchers with regard to open access publishing. The following sections will dissect the dimensions listed above and explore how they may either drive or inhibit Open Access adoption.

3.4.1. Social influence and affiliation

Zuiderwijk, Shinde & Jeng (forthcoming) denoted social influence as an important factor for open data adoption by researchers. Various literary sources enhance the claim social influence may either drive or inhibit open access adoption as well (Dulle et al., 2010; Klang et al., 2008; Swan, 2006). Swan (2006) cultivates an overview of researchers' views on open access and states 'communicate results to peers' as strongest driver for the use of open access. Knowledge sharing with one's peer community is rendered as top priority to researchers. Research culture shows potential to drive open access adoption. However, a study by Klang et al. (2008) resulted in finding that adoption by individuals does depend on group behaviour. That is, if a researcher perceives his or her peers having access to all renowned journals, one is not incentivized to adopt open access means.
Dulle et al. (2011) emphasizes the role of social influence by finding over 60 percent of researchers consider the attitude of close colleagues towards open access as being important. This percentage grows even larger when considering how leading researchers from the interviewee's research discipline perceive open access. The likelihood to employ self-archiving, i.e. green road open access, has been found to be dependent on social factors as well (Kim, 2011; Lwoga & Questier, 2014). That is, one is more (or less) likely to self-archive if one's peers decide to (not) do so as well.

3.4.2. Facilitating conditions

Facilitating conditions are defined as the degree to which one believes an organisational and technical infrastructure exists as to support the use of a certain system (Venkatesh et al, 2003). More specifically, it entails how researchers perceive technical and legal support from their institution, as well as the availability and trustworthiness of available platforms. We add to this definition by including the additional effort to engage in open access. Björk (2013) strongly emphasizes on IT infrastructure and indicates the criticality of technical support to facilitate open access adoption. Nonetheless, Dulle et al. (2011) found around 50 percent of their focus group states they lack knowledge on how to publish their work in open access outlets or lack resources and guidance in how to do so. The dissemination of technical instructions or the lack thereof significantly affect the willingness of researchers to engage in open scholarly communication (Klang et al., 2008).

The importance of legal support is widely acknowledged throughout literature (Kim, 2011; Klang et al, 2008; Björk, 2004; Forrester, 2015; Creaser, 2010). Kim (2011) states copyright concerns are prevalent throughout the academic community and is listed as a highly significant barrier to open access adoption. Publisher policies remain challenging to interpret and researchers often find themselves concerned with regard to copyright conditions and embargo periods (Creaser at al., 2010). Lwoga & Questier (2014) further underline this claim by including copyright concerns to be highly prevalent within academics employed at health science universities. As where ample opportunity for academic policy to mitigate legal concerns exists, current frameworks are often lacking (Klang et al., 2008). Researchers' needs do often not translate into the goals of a university's legal department and may oppose open access publishing rather than advocate its adoption.

Facilitating conditions also affect the perceived additional effort to publish through open access (Klang et al., 2008). Kim (2011) states that faculty members often resent additional activity that impairs their research and writing time. Hence, effort required to opt for open access rather than subscription-based journals is hard to justify, unless researchers see clear benefits. Researchers often find accessing open access repositories easier than disseminating information through open access (Dulle et al., 2011). Akin to open data sharing, effort is thus negatively correlated to open access adoption.

An example of facilitating means to publish open access are online repositories. Björk (2013) denotes subject and institutional repositories as separate categories in distinguishing between his barriers for open access adoption. Although both are closely coupled with green open access (i.e. self-archiving articles published in other journals), those concepts deserve further explanation. Subject repositories are research-discipline-specific databases, which contain publications for a certain field of study. PubMed and ArXiv are prime examples of well-known subject repositories, being dedicated to medical research and mathematics & physics respectively. Those repositories have surpassed the state of ordinary database and they are now regarded as an instrumental, journal-like tool in scholarly communication (Getz, 2005).

3.4.3. Business models

Due to the absence of subscriptions, the main source of income for membership-based journalism, open access requires different business models in order to be sustainable (Björk, 2004). A central mechanism for funding open access are Article Processing Charges (APCs) (Solomon & Björk, 2012). Essentially, APCs transfer the cost of funding from the subscriber to the authors, their employers or their funders. Since open access publications are free of charge to the reader by design, funding is incurred on the author's side (Pinfield, 2010). As where Dole et al. (2004) argue researchers are generally accepting towards publication charges and are even willing to meet higher financial requirements to make their research openly available, APCs as solitary business model faces an increasing volume of criticism.

Processing charges inhibit open access publishing and are generally appreciated negatively by researchers (Nariani & Fernandez, 2012). Allen (2005) specifies differences between disciplines in terms of general acceptance exist, as well as Article Processing Charges varying in size between journals.

Business models hold the potential to drive open access adoption if author-side financial burdens are mitigated (Pinfield, 2010; Nariani & Fernandez, 2013; Beasley, 2016). Pinfield (2010) provides policy-based measures to shape clear structure with regard to open access financing as to create a framework for researchers to operate in, including institutional arrangement for OA funding. Furthermore, library support enhances the likelihood of researchers publishing through open access (Nariani & Fernandez, 2012). Finally, Beasley (2016) advocates a rework of APC concepts as to facilitate smaller institutions to engage in open access publishing, as well as incentivizing open access rather than establishing financial barriers to adoption.

Brands of open access

Various forms in which open science is currently being practiced exist. As research has deemed the means of open access to be largely dependent on researchers' preference (Kraus, 2014), cross-examining different implementations of open access is rendered out of scope for this research. Nonetheless, a concise description of widespread alternatives are included within this section for completeness, terminology introduced here may also be referred throughout subsequent sections of this report.

Within open access, a ubiquitously adopted distinction is often made between green open access and gold open access. The former refers to self-archiving by researchers. That is, if allowed by their institution or journal, researchers submit their publications for free online publishing using a repository of choice. The latter refers to open access journals which, rather than a pay-to-read business model, allow their publications to be freely accessible by readers (Guédon, 2004).

3.4.4. Academic recognition

Career advancement and personal prestige are important drivers for researchers to publish their work (Swan, 2006). Researchers' productivity and resulting publications are highly associated with research grant and career success (Horta & Santos, 2016). Those drivers for publishing persist in the realm of open access and scholars are incrementally more likely to adopt open access if academic reward may be earned by doing so (Lwoga & Questier, 2014; Kim, 2011; Björk (2013).

Mercieca & Macauley (2008) state open access journals are often perceived as having lower impact rating and prestige. Impactfulness and the prestige of journals are heavily linked to academic recognition and low perceived scores for those dimensions inhibits open access adoption (Mercieca & Macauley, 2008). Academic reward systems, including tenure and promotion, often heavily rely on publication quality (Kim, 2011). Contrarily, Borgman (2007) advocates academic policy incorporating rewards for the usage of new forms of scholarly communication, such as open access. Nariani & Fernandez (2012) further underlines the importance of open access acceptance by university policy and grant councils. Reformed academic reward systems may transform the barriers posed by current frameworks into a driving force behind adoption (Borgman, 2007).

3.4.5. Visibility

The impactfulness of research largely depends on the exposure it gains through publishing. In a study on the impact of publishing during the PhD phase of a researcher's career, Horta & Santos (2016) found faster career development was achieved by scholars acquiring citations and publications during their PhD. Furthermore, it has been widely acknowledged the number of citations are a strong indicator of how influential that piece of research is (Ebrahim et al., 2014). The h-index, a ubiquitously adopted metric for measuring researcher impactfulness, further stipulates citation numbers into an impact ranking as well (Bornmann, 2007).

Open access offers auxiliary visibility benefits in comparison to subscription-based publishing (Gargouri et al., 2010; Creaser et al., 2010; Migheli & Ramello, 2014; Eysenbach, 2006). Through comparative analysis, it was found open access articles are cited earlier and are, on average, cited more often than non-open-access articles (Eysenbach, 2006). Those citation advantages are further acknowledged by Pinfield (2015), but it is stated impact of Open Access publishing supersedes the realm of journal impact rates alone. That is, paper-level metrics are to be included into quality assessment rather than defining visibility as mere citation rates. Garganti et al. (2010) advocates for increased self-archiving by researchers as to increase visibility and research impact. Scholars value visibility benefits as important towards open access adoption (Creaser et al., 2010; Nariani & Fernandez, 2012). Furthermore, academics expect open access means to perform better in terms of citations and visibility (Dulle et al, 2010).

Inhibitors related to visibility are largely absent from literature. That is, citation advantages are often assumed to exist if addressed within literature. However, Moed (2007) states advantages can largely be attributed to *faster* publishing under open access, than publishing it *freely*. Inversely, Wang et al. (2015) does find higher impact for Open Access publications for both citations and attentions. Furthermore, Lwoga & Questier (2014) obtain insufficient evidence for increased visibility as an encouraging factor for adopting open access scholarly communication. As where literature is generally supports the open access citation advantage, it is important to note antithetical studies exist.

3.4.6. Researcher’s background and personal drivers

Throughout literature, it is widely acknowledged age, nationality and seniority are all influential towards the likelihood of open data sharing, as well as open access adoption (Fecher, Friesike & Hebing, 2011; Schöpfel et al., 2016). Furthermore, this array of research also notes a discrepancy in the commonality to share data openly between disciplines. Therefore, requesting subjects to input this data seems paramount for drawing insightful conclusions. Despite this information not being included within the stated choice experiments, respondents will be asked to disclose their personal data as to infer demographic implications of experiment results.

Personal drivers and intrinsic motivations form a family of influential factors towards open data sharing as well (Zuiderwijk, Shinde & Jeng, forthcoming). As this category mainly concerns personality traits and beliefs, it would be significantly challenging to devise policy advice within this domain. One cannot simply strive to employ only people from a certain group of desirable personalities and similar beliefs; especially regarding the fact institutions consists of a large number of employees. Nonetheless, awareness on the influence of underlying values and personalities remains present throughout experimentation.

3.5. Open source

As where in literature, open science principles are often related to open access publishing and open data sharing, some have argued those concepts alone are merely a step towards open science – yet fail to achieve it. Hey & Payne (2015) raises concerns that true reproducibility, a transparency-related metric to assess whether experiments can be checked for truthfulness by peers, may only be accomplished if underlying code is shared as well.

The shared benefit of enhanced validation can be considered as the incipience of relating open science to open source, the term coined for the open sharing of code. Nonetheless, the array of literature on the synergy of open science and open source is rather sparse. Willinksy (2005) pinpoints this unacknowledged convergence, yet argues open initiatives share the underlying motive of liberating intellectual property to be freely accessible. For instance, as where open access distinguishes itself from journal-based publishing by absolution of subscription fees, open source software often takes a similar free-or-subscribe approach (Willinsky, 2005). Easterbrook (2014) acknowledges the positive effect of open source code sharing on reproducibility and code quality, yet argues sharing code openly is merely a first step. Barriers to sharing arise from, amongst others, 1) portability: code is optimized for specific platforms, 2) configurability: configuring models for runs may be difficult, 3) entrenchment: many layers of decision-making may go into the code (Easterbrook, 2014). Although these barriers are surmountable, Easterbrook (2014) divulges a Pareto distribution applies to the community generated by software shared openly. That is, only a small portion of software will attract a significant amount of attention, as where the majority will fail to do so. Considering the definition of open science provided in section 2.1.2, open source mainly categorizes itself into the realm of transparency – a claim backed by Lyon (2016).

3.5.1. Reproducibility & validation

A strand of research devoted to open source in the light of reproducibility stems from the realm of executable papers. Rather than containing the traditional array of knowledge, such as figures, tables and results, executable papers incorporate computational content as to allow readers to validate and explore experiments (Koop et al., 2011). Currently, the act of validation of any particular scientific effort is deemed difficult if the implementation of methods and source code are excluded (Kauppinen & Espindola, 2011). The usage of open source environments and therefore, open source code sharing, alleviate those hardships, as well as enabling executable papers as viable means of publishing (Kauppinen & Espindola, 2011).

3.5.2. Reusability & collaboration

Besides reproducibility advantages, open source catalyses collaboration and code reuse (Chen et al., 2014). With uniform, open source software such as R and Python, discrepancies between source codes seize to exist. Therefore, source code may be readily shared between and utilized by individual research bodies. Besides linguistic homogeneity, several proposals for open system communities, based on the principle of open source, have been documented (Price-Whelan et al., 2018; Lippert et al., 2019). Price-Whelan et al. (2018) describes the Astropy project, an open-source initiative for the astronomical community. Rather than a variety of disparate software solutions, it strives to create a uniformly accepted software package to perform astronomical calculations. Lippert et al. (2019) propose a similar community for the field of pharmacology. Although the community proposed by Lippert et al. (2019) supersedes the realm of open source – it also seeks to devise best practices and provide education – it can be considered another collective open source initiative. Sojer & Henkel (2010) provide quantitative evidence that the increased opportunities for code reuse can indeed be considered a driver behind the success of open source software development. Nonetheless, contradictory research suggests that an abundance of source code shared openly fails to attract an audience at all, being stored within repositories to be never reused (Beecher et al., 2008).

3.6. Scope

In conjunction with the research questions, the current array of research provides a wide variety of directions for this research to be aimed at. To assure comprehensibility, yet address an array of current knowledge gaps, it is of key importance to establish a crisp research scope.

The disparate aspects composing the hypernym that is open science, serve as a basis to devise such a research boundary. As where the importance of each component of open science is readily acknowledged, focus is laid on open data and open access. Not only is the array of drivers and inhibitors of open access adoption largely coherent with those observed for open data, they also extensively cover the various aspects of open science described in section 2.1.2. Furthermore, the behavioural nature of the main research question should be highly reflected in a holistic approach towards experimentation. That is, when examining the decision-making process of a respondent, one must seek to incorporate factors from each dimension he or she may render as relevant. Here, the grouped set of factors from open data and open access are considered appropriate to do so.

This is not to say the relevance of open source is disregarded. As where arguments for the inclusion of open data and open access were provided, two factors have led to excluding open source as a central topic. First, the relationship of open source to open science remains fuzzy. In contrary to the broadly investigated interplay of open data, open access and open science, open source has received far less attention. Not only does this result in ambiguity on the drivers and inhibitors behind open source adoption by researchers, it also inhibits finding an impetus for behavioural examination. Secondly, the current array of literature frames open source software as an auxiliary tool for transparency and reusability. Despite these being dimensions of open science, they only cover a fraction of its full definition.

3.7. Key Concepts

Due to the wide variety of approaches to open science present throughout literature, an overview of key concepts is provided here. Definitions stated within this section are not indefinitely assumed as ground truths, yet they serve as rationale for the purpose of this research.

- Open science: an umbrella term for a subset of aspects representing an open, accessible-to-everyone approach to science and the sharing and publication thereof. Also includes open courseware, open education and open software, yet those domains are not considered for this research.
- Open research: an aspect contained within open science. Open research is composed of open data, open access and open source. Open research is the main emphasis for this study and one may state open science equals open research for this particular case, as we forego other aspects of open science. Furthermore, OA, OD and OS are interchangeably used to abbreviate open access, open data and open source.
- Stated choice experiment: the methodology applied throughout experiment design, utilized to elicit preferences from participants. Stated choice questions are formulated according to two characteristics:
    - Attribute a driver or inhibitor for open access or open data, translated into a factor to be evaluated by subjects under varying levels.
    - Level: a value assigned to an attribute, to be varied over the course of disparate questions, evaluated by subjects as to elicit preferences.

3.8. Conclusion

This chapter delineates the research background. First, the definition of open science and its benefits are described. Here, we define open science as the *the dissemination of scientific knowledge that is as wide as possible, free of charge to all users, and accessible online*. Open science is a hypernym, consisting of multiple branches and schools. This research mainly concerns the pragmatic and measurement schools, indicating a focus on more efficient knowledge creation and alternate measurement systems for scientific impact.

In terms of the benefits of open science, literature strongly suggests an accelerated research cycle could be achieved by adopting open science principles. That is, open science allows for eliminating redundant data collection as well as offering opportunities for novel collaborations. Furthermore, visibility, transparency and reproducibility increases under widespread adoption of open science principles as well.

A plethora of drivers and inhibitors apply to each aspect of open science. Open data may be driven/inhibited by 1) policy and voluntariness 2) facilitating conditions 3) trust 4) social influence and affiliation 5) effort and 6) legislation and regulation. In terms of open access, the following dimensions apply: 1) social influence and affiliation 2) facilitating conditions 3) business models 4) recognition and 5) visibility. For open source, drivers and inhibitors revolve around 1) reproducibility and validation and 2) reusability and collaboration.

In order to scope the remainder of this research, a few considerations were made. Open access and open data occupy a central role during the experimental phase, due to their drivers and inhibitors strongly reflect the definition of open science opted for here. It is estimated the grouped set of attributes from open data and open access holistically cover factors influential to a respondents' decision-making process. Although open source exhibits a certain degree of relevance to the problem matter, its relationship to open science remains fuzzy. That is, the interplay of open access, open data and open science has been investigated broadly, yet open source has received far less attention. Furthermore, open source mainly pertains to the realm of transparency and reusability. Despite them being imperative pillars of open science, they merely comprise a fraction of its full definition. These considerations led to excluding open source from experimentation.

4. Experimental Design

4.1. Introduction

This section will outline the experiment design of this study. The general structure and components of the experimental setup are described. Within this chapter, both the demographic section and the stated-choice section of the experiment are outlined, along with their underlying design choices. Here, focus lies on devising the experimental framework and providing the impetus for data collection and analysis. Although no research question is answered within this section, it serves as an instrumental step in the research process and cannot be omitted in order to progress to experimentation. Please note that actual experiments and experiment questions are available through Appendix II.

4.2. Experiment goal

As stated within the research goals, one of the main contributions of this research is to deduct a prioritization of factors for open science principle adoption by researchers. The experiment goal can therefore be defined as deriving which drivers and inhibitors observed through literature are of key importance. Experiment results are expected to serve as impetus for conceptualizing a prioritization of factors, to be used as policy levers by universities.

4.3. Experiment structure

As to achieve the experiment goal, stated-choice experiments were found viable in order to do so. Nonetheless, certain drivers and inhibitors of Open Science adoption do not stem from a policy factors, but are rather dependent on individual background and experience. As researcher background and experience are not eligible for examination through stated-choice experiments, a bipartite experiment structure is opted for.

Ahead of stated-choice questions, participants were prompted for demographic information and their preliminary experience with open science concepts. By collecting information on a respondent's background, their familiarity with open science and their scores on demographic adoption factors can be determined. In conjunction with stated-choice-based preferences collected consecutively, correlation between background and prioritization may be distinguished.

4.4. Experiment Design

This section further describes the experiment design. For survey-based questions, the dimensions as well as their relevance are introduced. For the stated-choice experiment, the attributes and their corresponding levels are discussed, along with the questionnaire itself.

4.4.1. Demographic & respondent's background

The survey questions directed towards gaining insight into multiple, underlying factors for Open Science adoption. Components included within this part of the survey are based on those observed relevant within literature. A comprehensive overview is available through Table 3.1.

Information	Relevance
Age (1 question)	• May drive or inhibit willingness to adopt novel technology, i.e. Open Data/Open Access (Fecher, Friesike & Hebing, 2011; Schöpfel et al., 2016).
Country of residence/university affiliation (2 questions)	• Choice of population requires respondents to reside in The

	Netherlands and affiliation with at least one Dutch university • May serve to cross-reference results against data on open science for Dutch universities.
Current position/research discipline (2 questions)	• Research discipline is relevant towards open science adoption (Allen, 2005) because Open Science principle penetration differs between research disciplines • Current position/tenure may influence adoption (Kim, 2011)
Experience with open data (3 questions)	• Affinity with open data sharing and open data usage may affect choices made during stated-choice experiments. Those currently more involved with open data sharing, are likely to do so in the future. (Zuiderwijk, Shinde & Yeng, forthcoming)
Experience with open access (3 questions)	• Affinity with open access publishing and open data readership may affect choices made during stated-choice experiments. Those experienced with open access publishing are more likely to elect choices into that direction (Schöpfel et al., 2016).
Questions on grants & institutions	• Openness to accepting grants requiring Open Access publishing and/or Open Data sharing affect choices made during stated-choice experiments. If a respondent is currently required to publish through open access/share data openly by his/her grant, it is more likely he/she will lean towards picking choices into that direction (Nariani & Fernandez, 2012).

Table 3.1 Questions on demographics

4.4.2. Stated-choice experiment

To establish which factors are favoured by participants, a set of stated-choice questions is presented. A set of attributes and levels, i.e. attribute values, which may be altered between questions, is determined to perform this part of the experiment. Respondents are asked to elicit preference from two options. Despite commonalities between factors for Open Access and Open Data exist, it has been chosen to include them separately as to more precisely distinguish which attributes are of key importance. This section holds the definition of attributes, levels as well as experiment questions for stated-choice experiments on both Open Data and Open Access.

4.4.2.1. Stated-choice experiment design

Rather than opting for a full factorial experiment, which would entail including each possible combination of levels, it was opted to employ an efficient design to generate a variety of choice situations. Efficient designs attempt to minimize the standard error, employing priors to calibrate each attribute (Choicemetrics, 2018). Although indicatory data on priors could not be obtained from literature, the polarity of the priors could be accurately determined. Thus, priors used to generate the list of choice tasks were specified according to a small number (0.01) along with a polarity (+/-) . For instance, 'Effort', an attribute included for both open access and open data, was assigned negative polarity, as we do expect a higher level of effort to negatively impact adoption.

4.4.2.2. Efficient vs. orthogonal design & questionnaire design choices

Efficient & Orthogonal designs

Besides the efficient design opted for to design the stated-choice experiment for this study, alternative design methods exist as well. Orthogonal designs hold similar properties to efficient designs and greatly reduces the number of questions with regard to full factorial designs, yet does not allow the specification of priors (Molin, 2017). Nonetheless, efficient designs are able to reduce error margins and the number of questions required for an equal amount of respondents. Bliemer & Rose (2005) denote lower D-error as a measure for covariance between attribute levels shown to respondents. Hence, minimizing D-error leads to a more efficient design, reducing coherence betwee tasks (Rose & Bliemer, 2005). During experiment design in Ngene, similar D-error values were obtained for efficient designs with ten choice tasks as for orthogonal designs with twelve tasks. As prior polarity could be estimated accurately here, it was chosen to implement an efficient design. Appendix II holds a full overview of the syntax and designs obtained through Ngene.

Questionnaire design choices

In order to increase questionnaire comprehensibility and reduce tediousness, it was chosen to partition the choice tasks into two parts. That is, an individual respondent will only be required to answer a selection of choice tasks rather than the full set of twenty. For version A, a respondent is asked to indicate preferences for choice task 1-5 for both open data and open access. For version B, a respondent is subjected to choice task 6-10 for both open data and open access.

4.5. Stated choice experiment design: open data

4.5.1. Open data levels and attributes

Table 3.2 contains five attributes included for stated-choice questions regarding open data, as well as their levels. The attributes elected for inclusion reflect the drivers and inhibitors found through literature review. It is noted they correspond to the array of factors denoted in section 3.3. Only attributes susceptible to either policy or individual decision-making were included. For instance, respondents are unable to choose their personal background and hence, this could not be included as a dimension.

Attribute	Levels
Social Engagement (Open Data)	1. Peers are reluctant towards open research data and the sharing thereof 2. Peers are neutral towards open research data and may engage in Open Data 3. Peers actively share and re-use open research data
Effort (Open Data)	1. All support necessary to catalyse open research data sharing 2. Moderate IT & infrastructure support for open research data sharing 3. No IT & infrastructure support to help share research data
Recognition (Open Data)	1. Research institution (e.g. university) opposes openly sharing data and enforces policy that prohibits sharing

	2. Research institution (e.g. university) neither advocates nor opposes openly sharing research data through policies 3. Research institution (e.g. university) incentivizes openly sharing data
Control (Open Data)	1. The data repository holds full control over research data 2. Author retains rights to first use research data in a publication 3. Author remains in control of all rights regarding data management and usage
Data Quality (Open Data)	1. 50% or less confidence in data shared on the platform 2. 75% confidence in data shared on the platform 3. 100% confidence in data shared on the platform

Table 3.2 Open data attributes and levels

The levels chosen for each attribute were determined according to previous research within the field of open data. Since the application of stated-choice experiments to open data adoption is relatively novel, no similar experiment design were obtained. A multitude of sources served as a basis for attribute level design. Kim & Adler (2015) provided insight into the relevance of repositories and institutional pressure. Inspiration for the levels of effort and data quality was taken from here. Murray-Rust (2008) addressed copyright issues faced by researchers publishing open data and exemplified concerns on control by the target population. Gezelter (2015) offered insights into academic recognition for open data, as where social engagement was based on Zuiderwijk, Shinde & Yeng (forthcoming).

4.5.2. Open data choice tasks

The Ngene software package was employed to optimize error values and determine the required number of questions to assure design efficiency. The efficient design generated for stated-choice questions on open data is available through Appendix II. To illustrate the concept of stated choice tasks, a sample choice task is portrayed in Table 3.3. Under the conditions (i.e. levels) provided in the choice task, respondents are requested to elicit preference to either publish through open access or decide not to do so.

Attribute	Level
Social Engagement	Peers actively engage in the sharing and re-use of open research data
Effort	Your research institute offers moderate IT & infrastructure support for open data research sharing
Recognition	Your research institute (e.g. university) opposes sharing research data openly and enforces policy that limits sharing
Control	You retain the rights to first use your research data in a publication
Data Quality	You have 75% confidence in the data that is being shared on the data repository

Table 3.3 Sample open data task

4.6. Stated choice experiment design: open access

4.6.1. Open access attributes and levels

Table 3.4 contains five attributes included for stated-choice questions regarding Open Data, as well as their levels. Attribute relevance stems from literature review results and every dimension rendered as relevant is included.

Attribute	Level
Social Engagement (Open Access)	1. Peers are reluctant to publish using open access 2. Some peers do occasionally publish using open access 3. Peers do actively engage in open access publishing
Effort (Open Access)	1. Support by research institute and IT infrastructure (technical & legal) 2. Moderate support by research institute (both technical & legal) 3. No support by research institute and IT infrastructure (technical & legal)
Visibility (Open Access)	1. No difference in amount of citations between OA publishing compared to non-OA publishing 2. More citations for OA publishing than for non-OA publishing 3. Significantly more citations for OA publishing than for non-OA publishing
Academic recognition (Open Access)	1. OA publishing does not influence tenure and promotion 2. OA publishing somewhat influences tenure and promotion (recognized, but not ranked as highly as renowned subscription-based journals) 3. OA publishing strongly influences tenure and promotion (full recognition)
Publishing costs (Open Access)	1. No Article Publishing Costs 2. Low Article Publishing Costs (€1-1000) 3. Significant Article Publishing Costs (€1000+)

Table 3.4 Open access attributes and levels

The levels chosen for each attribute were determined according to previous research within the field of Open Access. Since the application of stated-choice experiments to Open Data adoption is relatively novel, no similar experiment design were obtained. Where possible, literature including level-based designs were consulted as reference (Dulle, Minish Majanja & Cloete, 2010; Creaser, 2010). Solomun & Björk (2012) provided an overview of Article Processing Charges for journals employing this business model and hence served as input for levelling the publishing costs attribute. Visibility is modelled after Eysenbach (2006) and abstains from assigning exact numbers to levels. Percentages and numbers were only included if they were expected to enhance comprehensibility or if they could be obtained accurately.

4.6.2. Open access questions

The Ngene software package was employed to optimize error values and determine the required number of questions to assure design efficiency. The efficient design generated for stated-choice questions on Open Access, is available through Appendix II. To illustrate the concept of stated choice tasks, a sample choice task is portrayed in Table 3.5. Under the conditions (i.e. levels) provided in the choice task, respondents are requested to elicit preference to either publish through open access or decide not to do so.

Attribute	Level
Social Engagement	Your peers actively engage in open access publishing
Effort	Your research institute fully supports open access publishing with IT infrastructure, counselling and legal support
Visibility	More citations for open access publishing than for non-open-access publishing
Recognition	Open access publishing somewhat influences tenure and promotion (recognized, but not ranked as highly as renowned subscription-based journals)
Publishing costs	Significant article publishing costs (€1000+)

Table 3.5 Open access sample task

4.7. Conclusion

This chapter outlines the experimental design of this study. A set of demographic questions serves as to relate experiment results to available data on open science principle adoption. This concerns age, research institution and current position, as well as an assessment of the level of familiarity with open science principles. Results from the literature review served as input for the attributes for both open data and open access. That is, each attribute reflects one or multiple factors observed throughout literature. As to reduce tediousness and to preclude correlation between choice tasks, a mathematical, efficient design is employed. As a result, the amount of choice tasks per respondent is limited to ten.

The readily completed experimental design enables the data collection phase to start. Over subsequent sections, the process of data collection and analysis is outlined.

5. Data Collection

5.1. Introduction

This chapter describes the data collection stage. Here, the tools utilized for survey distribution are outlined, as well as the strategy driving data collection. Besides indicating sample size requirements, this section holds a sample choice task, taken directly from Qualtrics, the experimental space of choice. This illustrates the interaction between the respondent and the experiment. Furthermore, one may learn about the survey distribution here, as well as the tools and communication channels utilized during distribution. Finally, the population requirements stipulated by the research demarcation are incorporated into distribution strategy. This section in itself does not answer a research question, yet precedes data analysis, which yields the relative importance of attributes specified in chapter 4.

5.2. Survey development & data collection

Ahead of gathering responses from the target population, it was first required to select a survey development tool and a means of collecting respondents' data. Qualtrics, an experimental space dedicated to the development and execution of survey-based studies offering various advantages, was elected for this purpose. Qualtrics is able to host experiments on its own servers and therefore stores responses within its projects. Hence, each response recorded by Qualtrics is stored and may be readily transferred to other devices to perform analysis operations. Qualtrics adheres to GDPR standards and is compliant with privacy laws enforced by the European Union. Furthermore, participants may access the survey through a common, direct link, precluding the necessity to devise other means of experiment sharing. Figure 4.1 depicts a sample choice task as shown to the respondent in Qualtrics.

Social Engagement	Peers are reluctant towards open research data and the sharing thereof
Effort	Your research institute provides full IT & infrastructure support to allow you to share your research data openly
Recognition	Your research institute (e.g. university) incentivizes sharing research data openly
Control	You retain the rights to first use your research data in a publication
Data Quality	You have 100% confidence in the data that is being shared on the data repository

Would you share your data openly on the platform specified above, given those circumstances?

Yes, I would openly share my data	No, I would not openly share my data

Figure 5.1. Sample open data choice task (from Qualtrics)

5.3. Data collection strategy

Due to the goals of this research, it is strictly required to assemble a diverse respondent population. This is required to, not only, reduce homogeneity in terms of researchers' background, but also allow for the observation of differences between disciplines in terms of open science principle adoption. In order to do so, the experiment was distributed to researchers from a variety of research disciplines. In section 4.3, specifically targeted research disciplines are listed. A mixed approach, split between directly targeting researchers through personal e-mail, as well as more general communication methods such as posts on Twitter were employed to ensure a broad audience was reached. The time span in which responses were collected ranged from June 9 until July 6.

5.3.1. Sample size requirements

In order to infer statistically significant results from data analysis, a certain sample size is required. Rose & Bliemer (2013, p. 1024) describes a general rule of thumb in assessing the number of samples necessary for the significance of stated choice experiments. The following equation applies.

$$N \geq 500 \cdot \frac{L^{max}}{J \cdot S}$$

where L^{max} is the largest number of levels for any of the attributes, J is the number of alternatives and S is the number of choice tasks, N constitutes the sample size. In case of the experiment described in Chapter 3, this would entail a sample size of 150 ($L^{max} = 3$, $J = 2$, $S = 5$). Therefore, this is the desired number of respondents to obtain statistically significant results from data analysis.

5.4. Survey distribution

In order to gather responses from the target population, a manifold of means was employed. Each type of communication channel tapped for data collection is described in this section.

5.4.1. Direct distribution

Compelling insights into open science adoption for individual research disciplines may only be obtained if sufficient responses from a certain research discipline are obtained. COVID-19 posed several restrictions on survey distribution. Closure of academic buildings and restrictions on physical encounters enforced communication to occur fully digitally. In order to do so, a group of researchers from certain fields of research were directly contacted through e-mail, as to gather an adequate number of responses for data analysis. Over the data collection time span, batches of personal e-mails were sent on at least 20 days, with an increasing amount of messages per session as time progressed. It is assumed approximately 75 e-mails were sent per session. Table 4.1 provides an overview of data collection-related numbers and figures.

Amount of personal e-mails (estimate)	Responses (from Qualtrics)	Response rate
1500	91*	6,1%
* includes responses from other sources, actual response rate is lower		

Table 5.1. Direct survey distribution in numbers

The following disciplines were targeted specifically through e-mail, as selected from the list of disciplines in Appendix I:

- Technology (including Computer Science/Mathematics)
- Sociology
- Biology
- Economics

This distinction is based on the taxonomy that science consists of four branches: physical sciences (physics, technology), biological sciences, psychological sciences and formal sciences such as mathematics (McGinn, 2012). Respondents from each branch of science were targeted as to obtain comparative data for cross-disciplinary analysis. In order to acquire responses from each domain, academics from the following institutions and faculties were explicitly targeted:

- Delft University of Technology: Computer Science, Technology, Policy & Management
- University of Groningen (RUG): Faculty of Science and Engineering (including biology)
- Erasmus University Rotterdam: Faculty of Sociology, Faculty of Economics
- University of Leiden: Faculty of Sociology

Direct distribution was operationalized through consulting researchers' pages hosted by Dutch universities. In doing so, their research discipline as well as their contact details were readily obtained. Figure 4.2 holds the template modified to address researchers personally.

Dear mr./ms. <LAST NAME>,

Open Science is often regarded as the future of science. Yet, far from every instance of scholarly communication is published through **Open Access**. Despite the rise of open data platforms, research data is not being shared openly either. This goes without saying that adoption exhibits large variations across research disciplines.

Why does this occur? Which factors are important for researchers in order to adopt open science principles? My thesis, conducted as part of my MSc Engineering & Policy Analysis at TU Delft, attempts to answer those questions.

In order to gain insight into **researchers' preferences**, I am currently conducting a **survey**. I am curious to find out your preferences with regard to Open Access publishing and openly sharing your research data. Therefore, I would like to invite you to participate. Given your background in the field of <RESEARCH DISCIPLINE>, you could allow for a better understanding of what the drivers and inhibitors are within your research discipline.

Would you help me out and complete my survey? It is readily available through the link below and takes about **10 minutes** to complete.

https://tudelft.fra1.qualtrics.com/jfe/form/SV_85EOUIZ7xdJuB5r

Your cooperation would be much appreciated, as well as the knowledge you would contribute being valued highly.

Thanks in advance.

Kind regards,

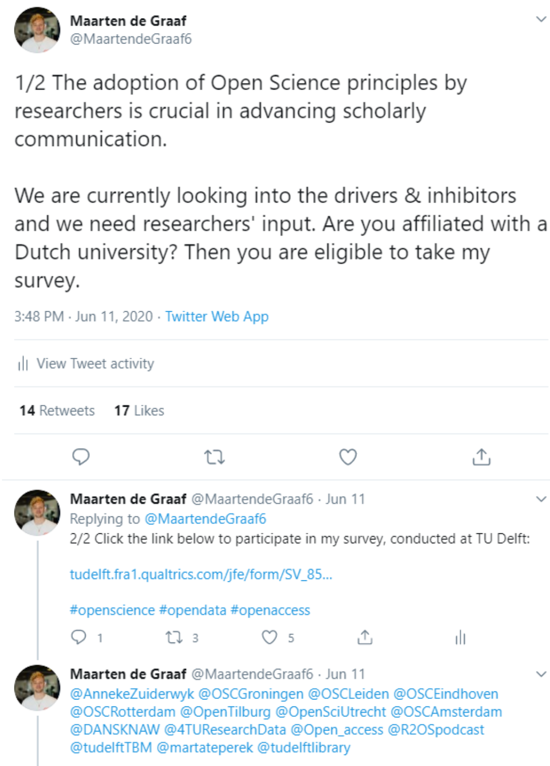
Maarten de Graaf
MSc Student Engineering & Policy Analysis
TU Delft

Figure 5.2. Template for direct survey distribution

5.4.2. Social Networks

A more general approach to data collection comes in the form of employing social networks. By means of sharing a post through the author's personal LinkedIn page, a multitude of responses could be obtained. Furthermore, Twitter served as a secondary social network to sensitize survey participation by the target audience. With the employment of several subject-related hashtags, a broad audience was amassed. Figure 4.3 shows an exemplifying thread shared on social media during data collection

Besides posts on Twitter and LinkedIn, experiment was included in a variety of newsletter by academic bodies. An item on the experiment was included in both the bi-weekly newsletter of the faculty of Technology, Policy & Management at Delft University of Technology, as well as the internal communications channel utilized by data stewards from this institute.



Figuur 5.3. Twitter thread shared during data collection

5.5. Conclusion

A sturdy data collection strategy ensures data set requirements are met and subsequent analysis will hold insightful results. A dual approach to data collection, including direct distribution and distribution through social networks, enables reaching the respondent group and collecting sufficient responses.

Respondent group heterogeneity is assured through explicitly targeting four core groups, based on the taxonomy of science into 1) technology 2) sociology 3) biology and 4) economics. Target groups include members from the following Dutch universities (departments between brackets): Delft University of Technology (Computer Science, Technology, Policy & Management), University of Groningen (Faculty of Science and Engineering, Biology), Erasmus University Rotterdam (Sociology, Economics), University of Leiden (Sociology).

6. Data Analysis

6.1. Introduction

The stage of data analysis succeeds data collection. This chapter outlines the most critical findings. First, a demographic analysis will provide insights into the composition of the respondent group. Demographics serve as a framework in which stated choice experiments can be examined and discussed. Subsequently, preferences will be elicited by analysing stated choice experiment results. Finally, results will be discussed and corresponding policy implication will be outlined. By means of analysing stated choice experiment results, **research question 2** will be answered:

What is the relative importance of factors that influence open science adoption?

Over the course of the analysis, demographic filters are introduced as to measure the interaction effect between disciplines and open science principle adoption, which addresses **research question 3**:

How does the relative importance of attributes vary across research disciplines?

Research question 2 and 3 stipulate a set of policy implications for a variety of decision makers within the policy arena. Section 6.6 contains an overview of these implications per group of decision makers. Thereby, it answers **research question 4**:

What are the policy implications of the attributes' relative importance?

6.2. General remarks on the dataset

Despite the extensive effort allocated to gathering responses, as well as the variety of means being employed, the response rate remained rather limited. After data cleaning and the elimination of incomplete responses, the dataset comprised of 91 entries. It is noted this is below the desired number of responses calculated in section 4.3.1. Orme (2019) states a lack of respondents increases standard errors and induces additional sample error during analysis. That is, the sample collected from respondents may deviate from the underlying population, which may lead to overestimating or underestimating the relative importance of factors (Orme, 2019; Rose & Bliemer, 2013). Nonetheless, given the time available for this research, it was decided to terminate the data collection phase and proceed to data analysis.

6.3. Demographics, involvement with open science principles & grants

Ahead of prompting respondents for their preferences regarding open data and open choice, numerous demographic questions were asked. Moreover, respondents were requested to indicate their involvement with open science principles and their likelihood to adhere to grant requirements with regard to open data sharing and open access publishing.

6.3.1. Seniority

Amongst others, Zhu (2017) observed significant differences in OA adoption with regard to seniority and age. Figure 5.1 represents the distribution of positions within the respondent group. PhD candidates and full professors are prevalent throughout the experiment population. This poses an interesting dynamic, since seniority positively correlates to open science principle adoption (Schöpfel et al., 2016; Zhu, 2017). Zhu (2017) states professors are more than twice as likely to deposit research articles using open access in comparison to researchers in training. Although stated choice experiments do not measure absolute importance or adoption, but merely the relative importance of factors, these claims will neither be confirmed nor refused here. However, with a high prevalence of senior researchers within the respondent group, it is expected factors related to benefits of open science (i.e. visibility, recognition) will prove as relatively important. Despite a tendency towards senior member of the Dutch research community, heterogeneity in terms of current position stems from Figure 6.1, which exhibits a mixture of different levels of seniority.

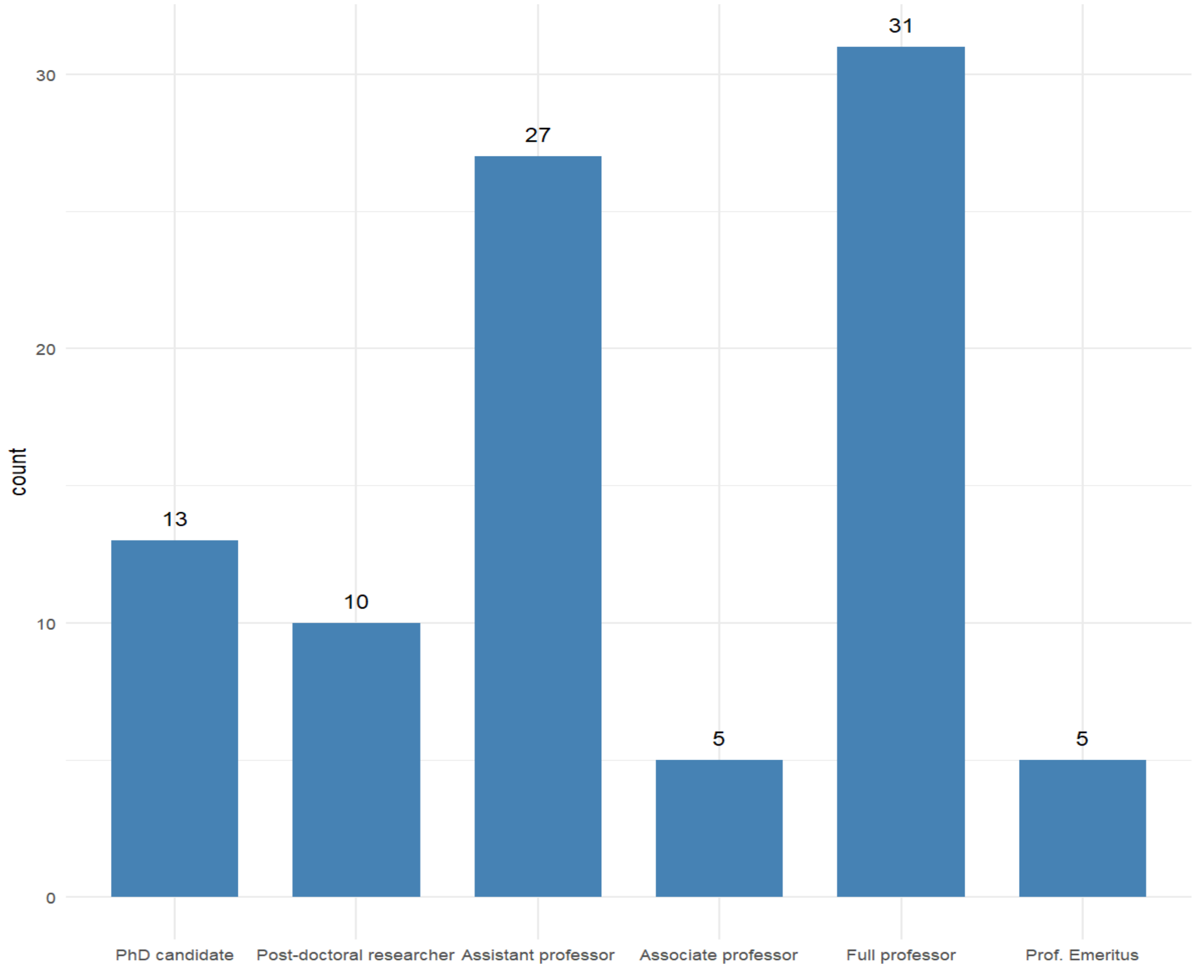


Figure 6.1. Current position of respondents

6.3.2. Research discipline & research institutions

Besides seniority, research discipline and research institute constitute a set of demographic factors gathered during data collection. Since institutional policy and research discipline were found influential with regard to the behaviour of academics in open science principle adoption, experiment results must be framed accordingly. Figure6.2 displays the institutional composition of the respondent group. It can be seen the majority of respondents is affiliated with universities that were explicitly targeted through direct distribution. Therefore, it is expected attitudes towards open science principles are coherent with policy governed by those institutions. This implies the following:

- TU Delft actively encourages open access through policy, stating objectives in a dedicated, strategic plan for open science (Haslinger, 2019). University policy stipulates that researchers must deposit an open access version of their publication in the TU Delft Repository. Furthermore, TU Delft Library hosts a plethora of agreements with publishers in order to mitigate publication costs, as well as an open access fund to finance open access publications (Open Access publishing, n.d.). In terms of open data, TU Delft hosts 4TU.ResearchData, an initiative to share data collected within its ranks openly. Respondents affiliated with TU Delft constitute a major part of the data set. Hence, it is expected data analysis reflects characteristics stipulated by TU Delft's open science governance. In case of open access, article publishing costs should be of lesser impact due to mitigation systems, as well as lower effort reflecting TU Delft's policy to facilitate researchers to engage in open access publishing. With regard to open data, the availability of 4TU.ResearchData encourages trust and data quality, as well as lowering effort.

- University of Groningen employs several means to facilitate researchers to engage in open science. In terms of open access, financial support policy aids researchers in funding open access publishing (Publishing Open Access: Open Access Discount, n.d.). Auxiliary tools are not available to researchers at University of Groningen, nor is open access publishing compulsory. In terms of open data, no supportive frameworks seem to exist. University policy specifies research data should comply with FAIR regulations, yet it abstains from proposing means to do so (Research Data Policy, 2015). Respondents from this group experience less support and face larger barriers towards open science principle adoption. Their research institution does not aid them with regard to open data, solely offering moderate support for open access as well. Therefore, attributes such as recognition and effort could see increasing relevancy, as researchers may expect their research institution to facilitate their step towards open science.
- Erasmus University Rotterdam (EUR) stipulates open access publishing by its researchers in the form of self-archiving (Publishing in Open Access, n.d.). Hosting its own repository, RePub, researchers are at least obliged to conform to self-archiving their work. For funding, EUR policy refers to national funding options. In terms of open data, the EUR does not employ transparent policy and merely hosts an online information page. As where open access policy encourages self-archiving, absence of open data policy at EUR implies a lower awareness of open data principles. Therefore, overall adoption, as well as the perceived benefits of open data are expected to be less prevalent within this group and generally affect results in this manner.
- Leiden University hosts an institutional repository and several agreements with publishers to reduce article publishing costs (Open Access – Leiden University, n.d.). Therefore, a variety of support systems may be readily accessed by researchers at this institution. With regard to open data, Leiden University does not have a strong, publicly available mandate on its policy. Although several documents on data management, which mention options for sharing and publishing data, are available, no apparent support system seems to exist. It is expected this leads to a lower prevalence of open data awareness, as well as different expectancies in terms of data quality, effort and recognition of open data sharing.

Analysis results are framed and discussed according to these observations.

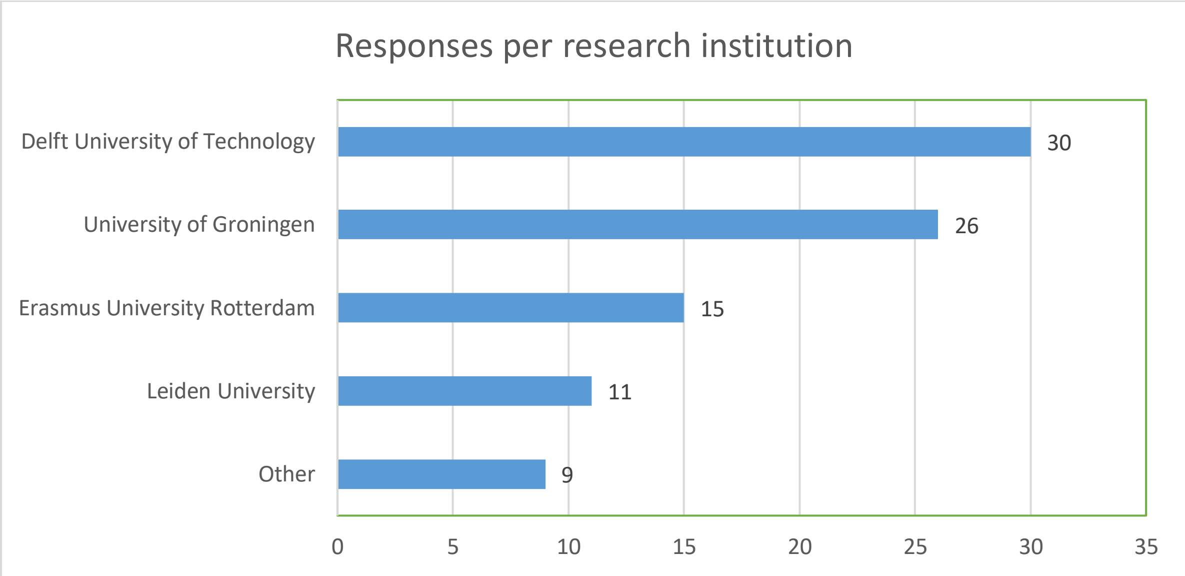


Figure 6.2 Responses per research institution

Figure 6.3 illustrates the responses per research discipline. Since this research is partially aimed at understanding the disciplinary differences of open science principles adoption, it was attempted to compose a heterogenetic population in terms of research discipline. Although Figure 5.3 implies this goal was achieved, the responses achieved per individual scientific domain are insufficiently sized in order to infer disciplince-specific behavioral patterns. Only if categorized into their superseding pillars of science (as described in 5.3.1), significant results may be obtained. Furthermore, the composition of research disciplines serves as a basis for interpreting stated choice results, as well as allowing for relating experiment results to the current array of literature on discipline-specific open science adoption. A few implications of the disciplinary compostions of the data set:

- Sociology-related disciplines constitute a significant part of the respondent group. Fry et al. (2010) observes a favourable stance towards open repositories in Social Science in Humanities, as well as a generally positive approach towards self-archiving. Therefore, it is expected social scientists exhibit a positive interaction effect with open science principle adoption. Similar observations hold for technology.
- Fry et al. (2010) observes contrary effects for biology-related disciplines, with a lower acceptance for self-archiving. However, Schöpfel et al. (2016) adds biologists are more receptive towards the payment of article processing costs. Hence, it is expected this attribute weighs less for this discipline.

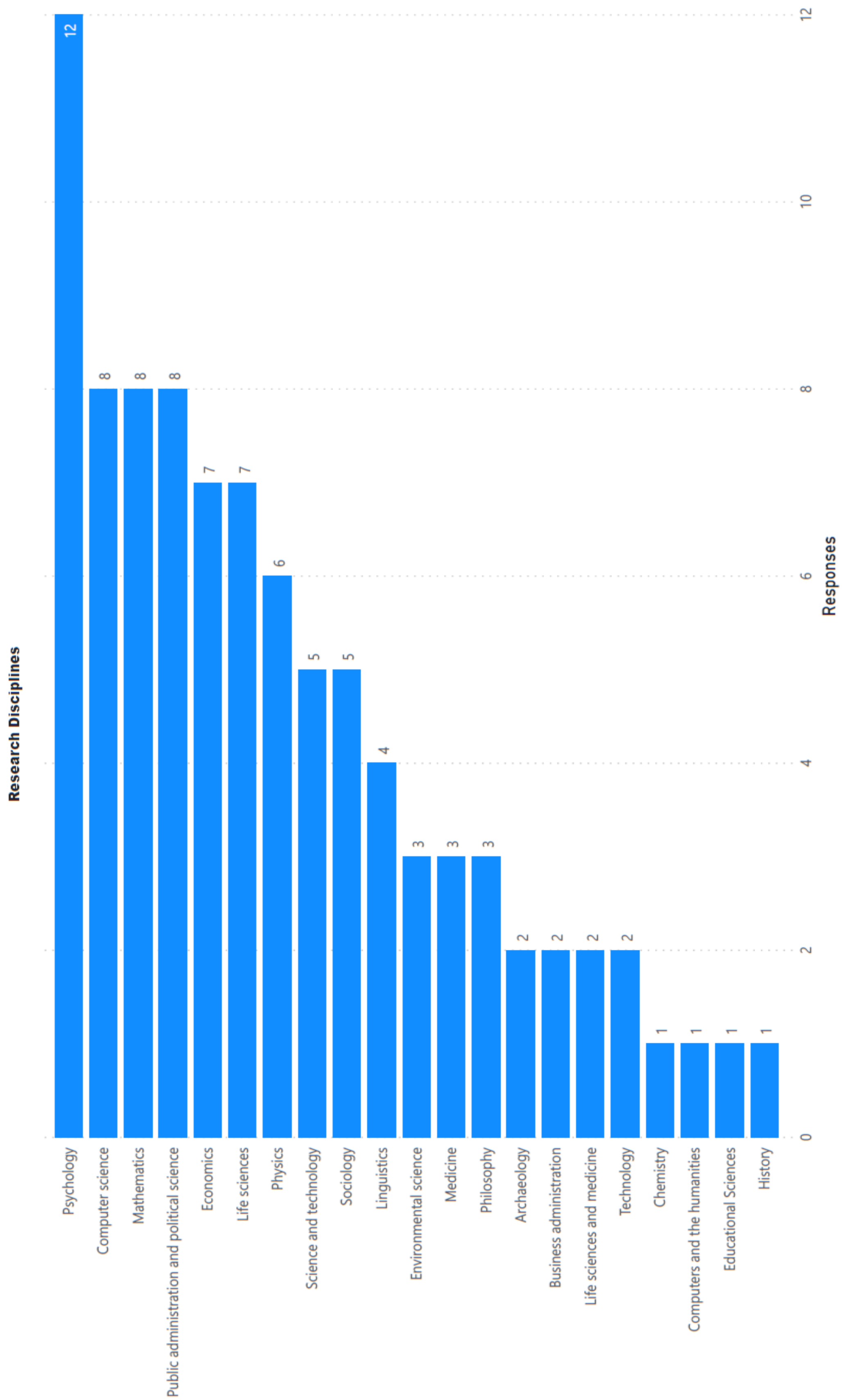


Figure 6.3 Responses per research discipline

6.3.3. Involvement with open science principles

The current affiliation with open science principles is evaluated through multiple questions. The involvement with open access, open data and open source were explored with regard to using resources shared through open science and resources contributed through open science. Participants were asked to assign a score between one (highly uninvolved) and ten (highly involved) based on their own involvement with open science. It is noted participants were first asked to specify whether they had ever engaged in open science. If not, questions on scoring their involvement were skipped. Table 5.1 provides descriptive statistics for each dimension, as well as involvement for each dimension. Furthermore, Figure 5.3 indicates the percentage of research published through open access by the experiment population.

	Mean	Median	SD	1st Quartile	3rd Quartile	% respondents involved
Open data (publication)	7.64	7	2.08	6.750	10	50,5%
Open data (usage)	6.96	7	2.01	6	8	Involved: 50% Unsure: 5%
Open access (usage)	7.48	8	2.07	7	9	96,7%
Open source (publication)	7.02	7	2.14	6	8	Involved: 36,2% Unsure: 8,8%
Open source (usage)	7.56	7.5	1.79	6	9	Involved: 53,8% Unsure: 4,3%

Table 6.1: Descriptive statistics open science principle involvement

Two key observations about the experiment population can be inferred from Table 5.1:

1. Respondents that had already engaged in either dimension of open science, perceived their current involvement as rather strong. With mean values at around 7-7.5, participants consider themselves engaging in open science often. Therefore, once aware of open science principles, respondents are likely to adopt them and re-employ them during future projects.

2. Open access usage dominates all other dimensions in terms of involvement. Only a small minority of respondents indicated to have never consulted research published through open access. Contradictory percentages were found for open source and open data. As where both the usage and contribution of open data is prevalent throughout 50% of the population, open source code contributions is heavily lagging behind. As final remarks provided by a multitude of respondents stated, datasets encapsulated within their research that involve third party data, are often restricted from open data sharing. Despite the willingness of researchers to share their data, the actual act of sharing is inhibited by third party contracts. From the low number of involvement in open source code sharing, it can be seen adoption is rather limited and that executable research is far from standard practice. It is however noted that researchers from certain disciplines are less inclined to apply coding within their research and therefore indicated to have never shared their code, despite there not being anything to share.

In contrary to the high prevalence of open access usage, the contribution of research published through open access is significantly lower. As can be seen from Figure 5.4, a significant of respondents indicated to publish only 0%-20% of their research through open access. Multiple remarks were made that, for specific disciplines or in general, the quality of open access journals is perceived significantly lower than subscription-based journals. Hence, respondents opted for subscription-based journals, disclosing that open access journals are merely a second option. Nonetheless, the largest subgroup of respondents indicated to publish 81%-100% through open access and they have (close to) fully adopted open access as their prime means of publishing. The large discrepancy between publishing through open access and consultation of open access sources exemplify the visibility benefits for open access publishing, yet indicate barriers to open access publishing have not seized to exist.

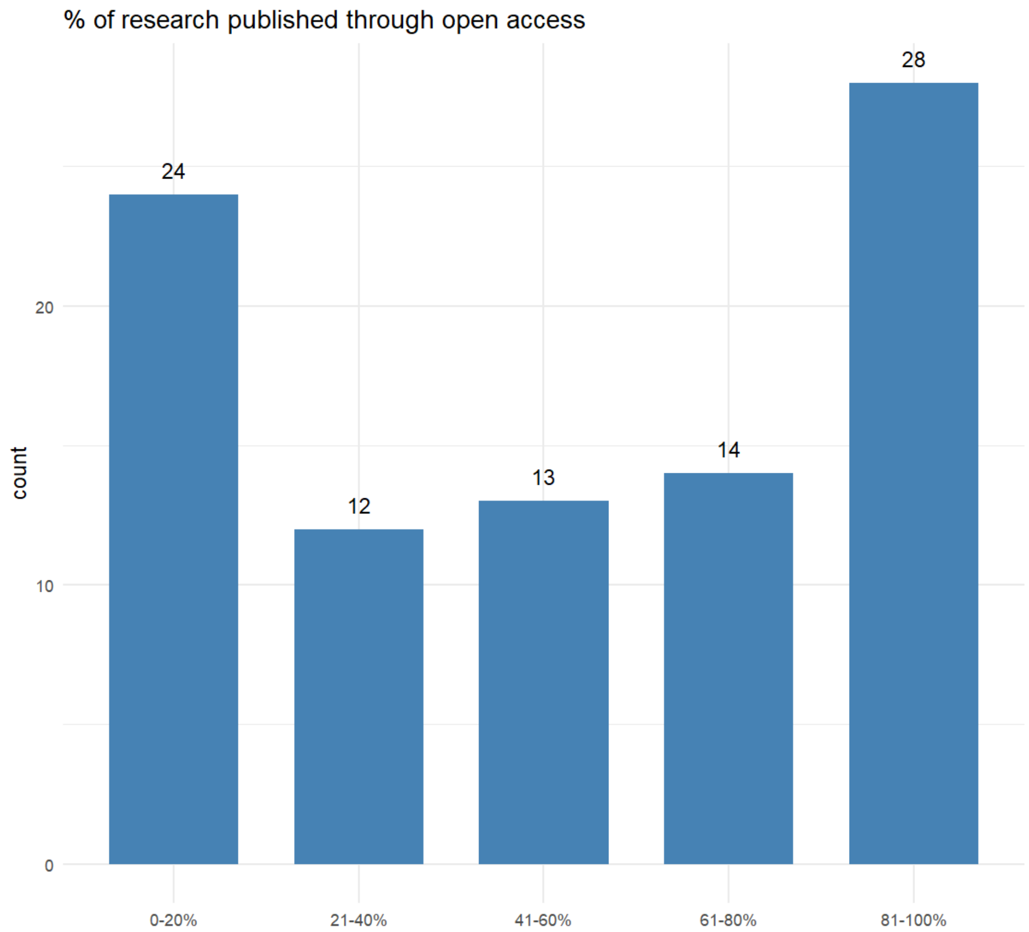


Figure 6.4. Percentage of research published through open access

6.3.4. Grants

During literature search, it was found grants were found influential in researchers' behaviour towards open science principle adoption. That is, if a grant requires an academic to adhere to certain open science principles, i.e. publishing research through open access or sharing research data openly, those taking this grant are more likely to employ such principles. It was asked from respondents to assign a score (ranging from 1 to 10) to the likelihood they would adhere to such requirements.

As can be seen from Table 5.2, the experiment population exhibits only a fair likelihood to comply with such regulations. Although open access requirements are more likely to be respected, statistical measures do not indicate high scores. This does not completely disqualify grants as a policy tool to encourage researchers to adopt open science principles, yet these results clearly indicate grants alone do not suffice as a policy lever alone.

	Mean	Median	SD	1st Quartile	3rd Quartile
Open data	6.09	6	1.16	6	7
Open access	6.46	7	0.89	6	7

Table 6.2 Grants & open science: likelihood to adhere to grant requirements on engaging in open science principles (rating 1-10)

6.4. Stated choice experiments

This section constitutes the analysis of the dataset gathered through stated choice questions. First, the analytical model of choice will be outlined, as well as several remarks on data manipulation. Subsequently, the experiment results for both open data and open access will both be analysed.

6.4.1. Choice model and data manipulation

Multiple remarks on data manipulation and the model elected for analysis apply. First, it was opted for to analyse the dataset through a MNL-type model. This type of model allows the analyst to infer trade-offs and observe the significance of each attribute included within a dataset (Train, 2009). Furthermore, relative importance may be derived from utility increases one gains by changing levels of attributes. These benefits cascade into the goals of this study, enabling the pinpointing of researchers' preferences. Train (2009) further elicits logit models such as MNL are capable of representing taste variation, yet cannot handle situations where unobserved factors correlate over time. Such correlation may be alleviated by employing Mixed Logit (ML) models, as they hold a variable encapsulating interference due to correlation. Due to time availability and the necessity to conjure a plethora of discipline-specific instances, it was decided not to include them for data analysis. This implies data results ought to be interpreted considering the MNL model may incorporate correlation into its estimated parameters.

In addition to model choice, several data cleaning and manipulation operations were required in order to prepare the dataset for analysis. Appendix III holds an excerpt of the source code and provides a link to an embedded Word-file with the full source code. Since initial set of responses merely includes respondent answers, it was required to inject the choice task table into the dataset through R. Apollo, the package employed to perform data analysis, stipulates strict dataset requirements, which demanded preparatory steps. Attribute levels that were defined as categorical values, were converted into numerical values through one hot encoding. With attribute levels being split into separate columns, level values were converted into binary, holding 1 for a choice task including this level and 0 for choice tasks not including this level. Besides this conversion, a variety of operations was performed on the data in order to trim, clean or filter specific demographic groups from the data.

6.4.2. Open data stated choice results: general model

In order to gain insights into researchers' preferences with regard to open data, it was first attempted to fit a general model to the dataset. That is, every response obtained was included for analysis, as well as each attribute. Considering the limited amount of responses gained, this was considered a necessary, exploratory step.

A multitude of statistical metrics reflects whether model outputs are to be considered significant, as well as the entire model performing appropriately. A variety of metrics relate to factor and model significance of factors. T-values and p-values indicate the significance of individual factors. For 5% statistical significance, it is required that 1) $t \gg 1.96$ or 2) $p < 0.005$ (Train, 2009). Furthermore, the Likelihood Ratio Test may be employed in order to assess the performance of the model itself. The Likelihood Ratio Statistic (LRS) is calculated as follows:

$$LRS = -2(LL_a - LL_b)$$

with LL_A and LL_B being initial and final likelihood values (Chorus, 2020). Schoonjans (2020) presents a a table to cross-reference LRS-scores against χ^2-values. If within range, the model achieves significance on a specified confidence interval. Over the course of analysis, these metrics guide the assessment of attribute and model robustness.

Furthermore, β-values are specified per attribute level, due to dummy-coding described in 5.3.1. For instance, β_{12} corresponds to attribute 1, level 2 or medium social engagement. B-values and their respective values should be interpreted in comparison to the omitted attribute level. That is, the estimate indicates how much utility one gains (loses) by selecting a non-omitted attribute over an omitted one. Finally, the sign of a β-value indicates a positive (negative) utility obtained by swapping alternatives, the size of a β-value indicates the utility loss (gain) obtained by swapping alternatives.

	Estimate	Std. Error	t-ratio	p-value
B_{12}	-0.883	1.0192	-0.09	0.931
B_{13}	0.2376	0.4320	0.55	0.582
B_{22}	0.8560	1.5089	0.57	0.570
B_{23}	-1.6933	0.9231	-1.83	0.067
B_{32}	1.4268	1.2166	1.17	0.241
B_{33}	3.1668	2.1940	1.44	0.149
B_{42}	-0.4856	1.4569	-0.33	0.739
B_{43}	0.3737	0.6278	0.60	0.552
B_5	-0.0065	0.0273	-0.24	0.811
Asc_A (common)	-0.2097	2.1265	-0.10	0.921
LRS	85.1274	**AIC**	529.59	
P^2	0.1431	**BIC**	570.21	
B_{1x} : social engagement	*B_{2x} : effort*	*B_{3x} : recognition*	*B_{4x} : control*	*B_{5x} : data quality*

Table 6.3: Stated choice experiment: open data MNL model

Table 6.3 holds the model output yielded by the MNL model for open data. A few remarks:

1. The model exhibits considerable significance issues. Most attributes were found to be insignificant, as p-values >> 0.005 and t-ratios << 1.96. The model outperforms the null-hypothesis with 1% significance (LRS = 85.1274 >> χ^2 = 27.877 (Schoonjans, 2020), yet ranks below average on other metrics. Rho-squared (0.1431) values between 0.2 and 0.4 indicate a good model fit, as where low AIC and BIC values demonstrates a model's focus on preventing information loss (Hauber et al., 2016). In general, low AIC and BIC values are thus preferred. This has not been achieved here. Hauber et al. (2016) state this indicates a loose model fit, meaning attributes may not accurately describe the decision-making process.
2. Due to high p-values, most attribute levels are to be considered insignificant in comparison to the omitted attribute level. For example, a perceived utility increase of 0.2376 is stated for switching to B_{13}, indicating the highest level of social engagement. Although this is estimated to increase the likelihood to adopt open data, it is estimated insignificant and therefore does not imply social engagement positively affects open data principle adoption. In contrary, an increase is rather insignificant. It is noted the polarity of insignificant attributes does not agree with expectations (Chorus, 2020). Therefore, they are left out of further discussion.
3. Factors related to the data platform, control and data quality, do not significantly alter behaviour under either increased control or perceived data quality. With t-ratios and p-

values far below significant values, it stems from the data researchers disregard data platform choices within their decision-making process.

4. B_{23} is the only factor approaching significance within the MNL model. With a sizable utility loss of 1.6933, an increase in effort to share research data openly can be considered an inhibitor of the likelihood to share data openly within the respondent group. From the t-ratio for $B_{33,}$ it may be inferred the polarity of recognition is in accordance with expectations. That is, a sharp increase in academic recognition positively affects utility and a researcher's likelihood to share research data openly.

6.4.3. Open data stated choice results: discipline-specific models

To infer discipline-specific behavioural patterns, it was sought to study interaction effects between different disciplines and open data principle adoption. For each discipline-specific model listed below, the utility function includes a general interaction effect, which assesses the general impact of a certain discipline on adoption. In each case, responses are assigned a binary value 1 if from the discipline under consideration, 0 if not. The discipline-specific models include interaction effects for specific variables as well, if they were found relevant by literature or in comparison to the general open data model.

Analysis of the biology model yielded no noteworthy results for open data; hence, it has been omitted from this section. The sparse amount of respondents from economic disciplines caused several model issues. Hence, they were excluded from disciplinary analysis as well.

6.4.3.1. Open data adoption: social science

Social science consists of a multitude of individual disciplines, namely psychology, sociology, linguistics, public administration and political science. Aggregating and encoding these disciplines into an overarching category served as input to the social science model.

	Estimate	Std. Error	t-ratio	p-value
B_{12}	-0.0610	1.0244	-0.06	0.953
B_{13}	0.2470	0.4276	0.58	0.564
B_{22}	0.7456	1.5809	0.47	0.637
B_{23}	-1.6383	0.9470	-1.73	0.084
B_{32}	1.3930	1.3190	1.06	0.291
B_{33}	3.0475	2.2269	1.37	0.171
B_{42}	-0.4634	1.4341	-0.32	0.747
B_{43}	0.3733	0.6170	0.61	0.545
B_5	-0.0049	0.02776	-0.18	0.859
B_{Disc}	0.5560	0.2357	2.36	0.018
Asc_A (common)	1.1196	0.8139	1.38	0.169
LRS	90,7946	**AIC**	525.93	
P^2	0.1527	**BIC**	570.6	
B_{1x} : social engagement	*B_{2x} : effort*	*B_{3x} : recognition*	*B_{4x} : control*	*B_{5x} : data quality*

Table 6.4. Stated choice experiment: open data for social science

Table 6.4 holds the model output resulting from including an interacting effect with social science. A few remarks:

1. The model exhibits similar significance issues to the model excluding the interaction effect for open science. Statements from 6.4.2 on significance remain intact. Similar remarks apply

to p-values; insignificance persists throughout the inclusion of interaction effects. Therefore, interaction effects for individual attributes were omitted.

2. The interaction effect between social science and open data adoption is significant. B_{Disc} represents the interaction between the utility obtained from belonging to a discipline contained within social science. As $t > 1.96$, β_{Disc} is significant within the 5% range. Since β_{Disc} = 0.5560, the hypothesis that social scientists are more likely to adopt to open data practices holds. It is noted this concerns a general effect, rather than an attribute-specific effect.

6.4.3.2. Open data adoption: technology

Technology aggregates the following disciplines: computer science, mathematics, physics, science & technology and technology itself.

	Estimate	Std.err.	t-ratio	p-value
B_{12}	-0.0926	1.0934	-0.08	0.933
B_{13}	0.2454	0.4586	0.54	0.593
B_{22}	0.8336	1.7809	0.47	0.640
B_{23}	-1.6883	1.0945	-1.54	0.123
B_{32}	1.4192	1.3197	1.08	0.282
B_{33}	3.1615	2.6326	1.20	0.230
B_{42}	-0.4957	1.6988	-0.29	0.770
B_{43}	0.3702	0.7168	0.52	0.606
B_5	-0.0063	0.0324	-0.20	0.845
B_{Disc}	-0.3735	0.2331	-1.60	0.109
ASC_A	-0.0976	2.4399	-0.04	0.968
LRS	87.7094	**AIC**	529.01	
P^2	0.1475	**BIC**	573.69	
B_{1x} : social engagement	*B_{2x} : effort*	*B_{3x} : recognition*	*B_{4x} : control*	*B_{5x} : data quality*

Table 6.5: Stated choice experiment: open data for technology

Table 6.5 holds the results for the model on open data, including the interaction effect for technology. Since the model exhibits similar scores for LRS, AIC, BIC and P^2, model fit and significance are similar to the general open data model. Furthermore, no disparate p-values and t-ratios were obtained. Hence, the only parameter of interest is $\beta_{Disc.}$ With a t-ratio < 1.60 and p > 0.005, absolute significance between technology as a discipline and open data adoption is left unproven. Nonetheless, the polarity from Table 6.5 holds, as the t-ratio approaches 1.96. Therefore, researchers from technology-related disciplines are less likely to share their datasets. Various respondents from this group disclosed a reluctance to share data openly, in fear of competing research groups utilizing it and eventually, outperform them. This could account for the negative relationship.

6.4.4. Open access stated choice results

Following the general model approach applied to analysing open data results, a similar approach was employed for open access data. Since performance metrics were already introduced in section 6.2.2, they are omitted here. One seeking to understand the analysis performed here is recommended to consult this section.

	Estimate	Std. Error	t-ratio	p-value
B_{12}	-0.6789	NaN	NaN	NaN
B_{13}	0.3272	0.3248	1.01	0.314

B_{22}	0.8818	0.9464	0.93	0.351
B_{23}	-1.2374	0.4286	-2.89	0.004
B_{32}	-2.31	1.2569	-1.84	0.066
B_{33}	-0.6783	NaN	NaN	NaN
B_{42}	2.4987	1.8258	1.37	0.171
B_{43}	1.2387	0.6958	1.78	0.075
B_{52}	0.8022	0.8026	1	0.318
B_{53}	-0.8332	0	-Inf	0
Asc_A (common)	1.1196	0.8139	1.38	0.169
LRS	194.0128	**AIC**	442.12	
P^2	0.3159	**BIC**	487.14	
B_{1x} : social engagement	*B_{2x} : effort*	*B_{3x} : visibility*	*B_{4x} : recognition*	*B_{5x} : publishing costs*

Table 6.6: Stated choice experiment: open access MNL model

Table 6.5 holds the model output yielded by the MNL model for open access. A few remarks:

1. Despite holding NaN values, the model is an appropriate fit for the model. LRS largely exceeds the null-hypothesis with 1% significance (LRS = 194.0128 >> χ^2 = 31.264, with df = 11, p = 0.001). Furthermore, rho-squared is in range of 0.2 and 0.4, indicating a good fit as well. Despite AIC and BIC being relatively high, other metrics indicate model viability.
2. Higher social engagement is proven insignificant with regard to behaviour. As both B_{12} and B_{13} exhibit insignificant/NaN scores, the relationship observed within literature is unconfirmed by stated choice experiments.
3. An increase in effort is strongly unfavourable with regard to open access adoption. With t >> 1.96 and p < 0.005, B_{23} indicates a large utility loss if effort were to increase. That is, the support provided by universities to incentivize open access publishing is influential with regard to researcher's behaviour.
4. An increase in publishing costs negatively affects the likelihood for researchers to publish through open access. In comparison to the lowest attribute level (no publishing costs), a negative utility is obtained from switching to the highest level of publishing costs.
5. Although insignificant within the t > 1.96 range, high recognition (B_{4x}) is preferred over low recognition. Both attribute levels exhibit utility gains, as well as a relatively high t-value. Therefore, the polarity can be assumed true and payoff from equally valuing open access publishing is therefore positive.
6. B_{32} holds a contradictory observation in comparison to expectations. One would expect an increase in visibility to lead to utility gain. Yet, a strong negative payoff is obtained for switching from the lowest attribute level (similar amount of citations for OA and subscription-based journals) and the middle attribute level (more citations for OA). Multiple comments from the respondent group stated they value subscription-based journals higher, regardless of the perceived amount of citations. This may have led to undervaluation of this dimension.

6.4.5. Open access stated choice results: discipline-specific models

To infer discipline-specific behavioural patterns, it was sought to study interaction effects between different disciplines and open access adoption. For each discipline-specific model listed below, the utility function includes a general interaction effect, which assesses the general impact of a certain discipline on adoption. In each case, responses are assigned a binary value 1 if from the discipline under consideration, 0 if not. The discipline-specific models include interaction effects for specific variables as well, if they were found relevant by literature or in comparison to the general open access model. That is, if a first iteration of the model, including the general interaction effects, yielded disparate results; attribute-specific interaction effects were introduced. Model fit and significance was similar to the general open access model for all discipline-specific models. Hence, LRS, BIC, AIC and P^2 values are omitted.

6.4.5.1. Open access adoption: social science

Social science consists of a multitude of individual disciplines, namely psychology, sociology, linguistics, public administration and political science. Aggregating and encoding these disciplines into an overarching category served as input to the social science model.

	Estimate	Std.err.	t-ratio	p-value
B_{12}	-0.7213	NaN	NaN	NaN
B_{13}	0.3630	0.2339	1.55	0.121
B_{22}	0.9193	0.9445	0.97	0.330
B_{23}	-1.2794	0.4642	-2.76	0.006
B_{32}	-2.4681	1.2769	-1.93	0.053
B_{33}	-0.7213	NaN	NaN	NaN
B_{42}	2.9273	1.8325	1.60	0.110
B_{43}	1.2881	0.6880	1.87	0.061
B_{52}	0.8862	0.8543	1.04	0.300
B_{53}	-0.8109	0.4310	-1.88	0.060
B_{Disc}	0.3272	0.2339	1.40	0.162
B_{Disc_42}	0.7255	0.4523	-1.60	0.109
Asc_A (common)	0.9594	0.8305	1.16	0.248
B_{1x} : social engagement	*B_{2x} : effort*	*B_{3x} : visibility*	*B_{4x} : recognition*	*B_{5x} : publishing costs*

Table 6.7 Stated choice experiment: open access for social science

Table 6.7 holds the model output for the open access model specific to social science. The following applies:

1. A first iteration resulted in finding disparate values for B_{42}. Hence, an interaction effect measure (B_{Disc_42}) was introduced. This beta-value measures the relative importance of academic recognition to social scientists.
2. Although the general interaction effect (B_{Disc}) is insignificant within the 5% range ($t < 1.96$, $p > 0.5$), it approaches significance. Therefore, it can be deducted the polarity of B_{Disc} is accurate. As a result, researchers from the field of social science are more likely to engage in open access publishing. This reinforces claims by Schöpfel et al. (2016), who states similar results from a study in France. Similarly, B_{Disc_42} exhibits insignificance, yet its polarity implies social scientists are driven by increased levels of academic recognition if faced with the decision to publish through open access.

6.4.5.2. Open access adoption: technology

Technology aggregates the following disciplines: computer science, mathematics, physics, science & technology and technology itself.

	Estimate	Std.err.	t-ratio	p-value
B_{12}	0.3254	NaN	NaN	NaN

B_{13}	0.8809	0.3330	0.98	0.328
B_{22}	-12.448	0.5661	1.56	0.120
B_{23}	-23.392	0.4249	-2.93	0.003
B_{32}	-0.6910	1.2010	-1.95	0.051
B_{33}	2.551	NaN	NaN	NaN
B_{42}	1.2659	1.3500	1.89	0.059
B_{43}	0.8266	0.4476	2.83	0.005
B_{52}	-0.8578	0.5900	1.40	0.161
B_{53}	-0.2464	0.4329	-1.98	0.048
B_{Disc}	1.1897	0.2598	-0.95	0.343
Asc_A (common)	0.3254	0.2895	4.11	0.000
B_{1x} : social engagement	*B_{2x} : effort*	*B_{3x} : visibility*	*B_{4x} : recognition*	*B_{5x} : publishing costs*

Table 6.8 Stated choice experiment: open access for technology

Table 6.8 holds results for the model with a general technology interaction effect included. Since both t-ratio and p-value are low, no interaction effects may be deducted from this simulation. Schöpfel (2016) indicates researchers from technology-related disciplines are generally more favourable with regard to open access, those claims are not confirmed here.

6.4.5.3. Open access adoption: biology

Biology aggregates biology, life sciences, medicine and life science & medicine into a disciplinary group. The introduction of B_{Disc} , a proxy for the interaction between adoption and biology, significance of recognition sharply increases in comparison to the general model. No significance between biology disciplines and open access adoption stems from the model, with t-ratio = 1.68 < 1.96 and p-value = 0.093 >> 0.005, no relationship may be inferred. Polarity inferred from the model results in finding that the biology discipline relates to open access adoption positively. Eger et al. (2016) claim biologists are more receptive towards article publishing costs. However, B_{52} and B_{53} exhibit similar significance and size for both the general model and the biology model. This indicate biologists neither obtain nor lose greater utility from increases in publishing costs. Hence, claims by Eger et al. (2016) are revoked here.

	Estimate	Std.err.	t-ratio	p-value
B_{12}	-0.6272	NaN	NaN	NaN
B_{13}	0.3254	0.3307	0.98	0.324
B_{22}	0.8809	0.6055	1.46	0.145
B_{23}	-1.2448	0.4615	-2.72	0.007
B_{32}	-2.2192	0.8431	-2.63	0.008
B_{33}	-0.6272	NaN	NaN	NaN
B_{42}	2.2980	1.3500	3.17	0.002
B_{43}	1.1408	0.4476	4.50	0.000
B_{52}	0.6986	0.5900	2.75	0.006
B_{53}	-0.7256	0.4329	-1.69	0.092
B_{Disc}	0.7836	0.2598	1.68	0.093
Asc_A (common)	1.0918	0.3746	2.92	0.004
B_{1x} : social engagement	*B_{2x} : effort*	*B_{3x} : visibility*	*B_{4x} : recognition*	*B_{5x} : publishing costs*

Table 6.9 Stated choice experiment: open access for biology

6.5. Discussion

A plethora of remarks applies to the results of the preceding analytical efforts. As where stated choice experiments for both open data and open access yielded novel insights, they should be framed according to overarching circumstances.

The model for open data is undeniably flawed in terms of significance. As both overarching metrics related to model fit and individual significance metrics exhibit imperfect values, the general viability of the model is impaired. Various factors should be taken into consideration given this issue. As noted before, the sample size of the dataset is rather small. Therefore, one may expect oscillations to occur and outliers to have a larger effect on model stability (Rose & Bliemer, 2013). A more comprehensive dataset would mitigate such obstacles and increase model fit altogether. Yet, this is not considered the prime explanation of the issues that arose. Hauber et al. (2016) states a critical pitfall of conditional logit models, namely preference heterogeneity. That is, logit models neglect differences in preferences across respondents. With such a discipline-heterogeneous experiment population, it is highly likely discrepancies in preferences with regard to open data sharing exist. In order to forego such disparate preferences, it was attempted to construct a discipline-specific model for open data, yet the sparse amount of data points heavily impeded significance, rendering it impossible to successfully distinguish disciplinary differences. Moreover, multiple respondents have indicated they are bound to third-party data contracts, prohibiting them from sharing data at all. As a result, they are rendered unable to share their data, regardless of the benefits they would gain from doing so. Besides, respondents stated that copyright constraints and the highly competitive environment they operate in impairs their likelihood to share their data openly. It is however noted the former may be alleviated by proper support – a factor encapsulated into the 'Effort' attribute. The prevalence of such restrictions within the group of respondents is likely to have affected the model, hampering its viability. One could argue the results obtained are somewhat in accordance with literature, since effort negatively correlates to the likelihood to share research data openly. Yet, a variety of preferences that were hypothesized as being true, were not observed as significant all. Although their relative strength remains unknown, one could also frame results obtained here as a rebuttal to their claimed importance. Various opportunities for the encouragement of open data exists, based on results obtained within this research. Three out of four institutions employing the majority of the respondent group does not actively support, facilitate or finance open data sharing. With effort prevailing as critical factor in open data adoption by researchers, research institutions find themselves at the core of incentivizing open data from inside out.

With regard to open access, significance proved less troublesome. Yet, a variety of factors should not be omitted from discussion. The exposure open access publishing has gained over recent years is clearly reflected within the respondent group, exhibiting a high number of contributions and usage of open access sources. Nonetheless, open access publishing is not ubiquitously accepted as the industry standard. A recurring comment during the experiment was that respondents valued the quality of subscription-based journals higher than open access journals, irrespective of their perceived benefits. Therefore, one could offer them an abundance of benefits from open access publishing; they would still proceed to publish in subscription-based journals. A pitfall of the list of attributes for open access is that they merely seek to represent journal quality, yet abstaining from including an attribute for quality. Although this dimension was omitted for the reason journal quality is impervious to policy or behaviour, respondents claimed the assumption that both journals are of equal quality to be untruthful. This is highly likely to have affected both the size of β-values within the model as well as attribute significance. Moreover, this may account for the contradictory, negative correlation found between an increase in visibility and likelihood to publish through open access. Policy frameworks to reduce publishing costs and reduce effort employed by TU Delft and

Erasmus University Rotterdam, two of the four institutions most prevalent within the respondent group, may therefore very well not adequately incentivize researchers to adopt open access. Although self-archiving is compulsory at these institutions, it does not convince scholars to adopt open access as their preferable means of scholarly communication. The heterogeneity infused within the experiment population is likely to have induced noise into the model. With qualitatively adequate open access journals not being prevalent within each research discipline, not every participant will possess equal means to adopt open access publishing as his/her prime channel of scholarly communication. Although two out of four research institutions that were most prevalent within the respondent group employ various means to facilitate open access publishing, effort and publishing costs prevailed as most significant factors for adoption.

Discipline-specific models yielded a variety of novel insights, yet posed a few interesting subjects for discussion. The impact claimed by the current array of research was not reciprocated by results obtained here. If interaction effects existed, their corresponding utility scores were limited. As where literature suggests researchers from the field of social science and technology are more receptive towards open access, proof for interaction only held for social science during data analysis. Furthermore, interaction between technology and open access adoption remained absent. Similarly, incthe willingness to commit to higher article publishing costs is not prevalent within biologists, more so than in general, revoking claims by Schöpfel (2016) on this matter. Discipline-specific research within the field of open data adoption is rather unseen throughout literature. Therefore, the negative interaction between technology and open data adoption sparks thought for further research. On a final note, significance problems apparent within the general models for open data and open access, traversed into the discipline-specific models as well. Due to sample size issues, erroneous results were obtained after the inclusion of attribute-specific interaction effects. Furthermore, responses from economic-related disciplines were significantly sparse and as a result, were fully excluded from analysis.

6.6. Policy implications

The learnings from data analysis can be percolated into a variety of policy recommendations. Although open science is often considered a hypernym that encapsulates the concepts of open data, open access and open source, it is not recommended for decision makers to form policy in such a manner. Despite the limitations posed by the analysis results, it is evident the pillars that constitute open science should be treated – and thus incentivized – individually. This section enumerates a number of policy implications, grouped by decision maker within the policy arena.

6.6.1. Research institutions

The following policy implications apply to research institutions:

1. *Research institutions should lower the bar for open science principle adoption by support, guidance and institutional repositories*

For both open data and open access, it was shown effort significantly inhibits the likelihood of researchers to adopt open science principles. Therefore, research institutions hold the opportunity to facilitate their members as to incentivize openness. The presence of IT infrastructure and (legal) support is instrumental in encouraging researchers to openly publish their work. It was found that three out of four institutions most prevalent within the respondent group do not actively engage in facilitating open data. As where research institutions cannot alleviate third party data restrictions, hosting an institutional data repository, offering guidance and counselling as to how to share data openly will significantly lower the bar for researchers to engage in open data. Institutional repositories are more prevalent for open access publishing, yet are equally important for open data adoption.

2. *Research institutions should form alliances with journals, financers and national initiatives to increase adoption*

Large-scale adoption calls for a holistic approach towards open science. Although research institutions occupy a central role with regard to open science, this study has shown external factors are of key importance as well. Respondents remain to question the quality of open access journals, whilst some of them lack open access means of scholarly communication within their field. As where behaviour with regard to open data remains largely unexplained, ample barriers to adoption surfaced during experimentation. It is not the role of research institutions to alter the nature of third party contracts or alleviate copyright issues, but an integrative effort between stakeholders could be spearheaded by research institutions nonetheless. Taking into consideration researchers are likely to adhere to open science-related requirements stipulated by grants, seeking the alliance of those issuing allowances seems a fruitful collaboration to explore.

3. *Encourage researchers to adopt open science principles through academic recognition*

In both cases, academic recognition for open science engagement has proven relevant for adoption. In contrary to the preceding recommendation, academic recognition is governed solely by research institutions themselves. With regard to open access, closing the gap between the appreciation of subscription-based journals and open access journals would be a strong step into the right direction. Not only will this increase the perceived quality of open access journals, career-driven individuals will feel incentivized to adopt open means of scholarly communication as well. Furthermore, egalitarian policy for the recognition of OA contributions may dilute the differences between journals when considering career advancement. In terms of open data, academic recognition proved less significant. However, researchers obtain a net positive effect from additional levels of recognition for open data sharing. Bilateral policy on recognizing both open access and open data for academic career advancement encourages openness and will, disregarding third party data restrictions, incentivize open data sharing amongst researchers.

4. *Raise awareness on financial policy and expand funding*

With academics burdened with the costs of open access publishing and APCs as the dominant business model, a new approach towards funding scholarly communication is required. During experimentation, a respondent remarked his research group had formed a shared monetary fund from which the expenses of open access publishing could be financed. Such community initiatives exemplify the need for university policy with regard to financing open science. Although various institutions hold financial agreements with journals or offer funding opportunities for open access publishing, the respondent group expressed hardships in this regard. This could either indicate awareness on current financial policy is lacking or prove current funding is underwhelming in comparison to the community's needs. The findings presented here indicate publishing costs are influential towards researcher's behaviour and the introduction of adequate financial frameworks will only serve to propel the prevalence of open access publishing further.

5. *Encourage open data adoption through university policy*

A surprising finding is the discrepancy between university policy on open data and the observed relevance of effort in open data adoption. Only two out of four universities most prevalent within the respondent group host institutional data repositories and actively promote open data sharing. As reducing effort is paramount for researchers to engage in open data, it is strongly advised for research institutions to engage in doing so. This could either be implemented through means mentioned before: academic recognition or IT and legal support, but through university policy, stipulating research must be shared always as well. Such policy is already in place for open access publishing at both TU Delft and Erasmus University Rotterdam. Therefore, it is recommended to extend this policy to the realm of open data. It is noted this brand of policy should incorporate third party restrictions to data.

6.6.2. Journals

Journals hold the opportunity to contribute to the policy arena in various ways, it is noted they only contribute to open access. The following recommendations apply:

1. *Aid in lowering the bar for open access publications with lower costs and support*

Journals hold the power to devise business models to cover their expenses. Although article processing costs are currently the dominant means of financing open access journals, other means of financing open access could be explored (Beasley, 2016). Agreements with universities, of which some are already in place for the institutions under investigation, may alleviate the financial burden experienced by researchers as well. Journals are strongly recommended to continue forging these alliances and shifting the costs away from individual researchers to research institutions or financing agencies. Although literature suggested some disciplines are more receptive to article publishing costs, these claims were revoked here. The power of journals in opening up science is exemplified by Elsevier's open access deal with VSNU, a national funding agency, and Dutch universities (Schoonen, 2020). By means of this agreement, members of Dutch universities can now freely access Elsevier's full journal portfolio, as well as unlimitedly publish articles through open access. Agreements such as these do not only severely lower the bar for researchers to adopt open science principles, it also surmounts the quality concerns respondents issued on open access journals. If additional renowned journals were to follow Elsevier's example, open access adoption would start to thrive.

As where effort reduction should start at research institutions themselves, journals are advised to address process-based barriers faced by researchers conjointly. This could concern means such as providing informative sessions in collaboration with universities, but also a thorough assessment of the current submission process.

2. *Focus on journal quality, not visibility*

The visibility advantages of open access publishing were found not to influence researcher behaviour in a positive manner. Therefore, open access journals should abstain from promoting visibility gains as primary benefit of open access publishing. It is recommended to attain an approach focused on journal quality. As comments made by respondents indicated, open access journals are considered to be of lower quality within a variety of research disciplines. The benefits offered by open access publishing will therefore never outweigh the disadvantage of publishing in a lower quality journal. For disciplines in which no eminent, high quality open access journal exists, it is encouraged for large-scale journals (i.e. DOAJ, ArXiv) to forge efforts in order to close this gap. Although it is noted perceived journal quality is a function of longevity as well, offering researchers a new publishing option in a renowned open access journal could alleviate these concerns. However, subscription-based journals hold status, based on their longer existence. This issue may only be surmounted by agreements akin to the Elsevier agreement.

6.6.3. Grants & Financers

Grants and financers occupy a central role towards financing open access. Besides financial capabilities, grants hold the auxiliary opportunity to stipulate open data and open access regulations within their agreements. Although it was found the likelihood of researchers to adhere to requirements on open access/open data by grant agreements, its role in incentivizing open science is not negligible. Researchers perceiving open access journals as inferior within their field of study, are less inclined to accept grants stipulating open access publishing. However, for those not subject to this notion, grants hold an effective policy lever.
Similar remarks apply to open data. Those not limited by third party data restrictions or faced with high competition, are more likely to adhere to open data requirements. Grants do not hold capabilities to, so to say, 'flip the board' on overarching phenomena such as data confidentiality and open access journal quality, but their influence serves as a significant policy lever. Within collaborative frameworks with research institutions and journals, grants and financers should focus on alleviating the financial burden for researchers.

6.6.4. Governments

On a national and international scale, governments hold a significant role in incentivizing the adoption of open science principles. Considering The Netherlands is subject to both the national government and the EU as a governing body, policy should be formed with respect to superseding jurisdiction. As where bodies such as research institutions and journals are unable to alleviate third party restrictions on data, the European Union could play a significant role in opening up data. Enforcing policy such as the General Data Protection Regulation, which controls the management of personal data, the European Union displays regulatory capabilities of installing policy within the field of research data.

Both national governments and the European Union are capable of forging (inter)national alliances between stakeholders. Given their sheer size and regulatory power, these institutions may constitute a driving force in closing agreements akin to the Elsevier agreement (Schoonen, 2020). Furthermore, funding national or European initiatives to finance open access, open data and auxiliary repositories lies at the core of effective policy formation by both the European Union and the national government of the Netherlands.

6.7. Conclusions

This section addresses the relative importance of factors for open access and open data adoption. In apparent order to structure this section accordingly, conclusions are separated into open data and open access.

6.7.1. Open data

Despite significance issues, a plethora of conclusions on open data adoption may be drawn. Mainly, this research provides a rebuttal to contemporary open data literature. As where various sources claim the importance of data control and perceived data quality, these did not significantly affect adoption within the respondent group. Furthermore, social engagement exhibits a loose effect on open data adoption by researchers, despite being apparent from literature. In contrary, increases in effort negatively influence the likelihood of researchers to share their research data openly.

Incorporating disciplinary interaction effects into the model yielded the observation that technology-based disciplines are less inclined to share their research data openly. Comments by respondents indicated this stems from competition between research groups. This was further exemplified by a flattened significance of ‘Effort’ as an attribute. Furthermore, social science exhibits a positive interaction effect with open research data sharing. Hence, social scientists are more inclined to share their data openly, despite noting they often manipulate restricted third party data within their research.

In terms of policy implications, a few recommendations apply. Research institutions are strongly advised to employ means to facilitate their staff to share research data openly. Superseding legal, IT and general support, research institutions should encourage sharing through university policy and maintaining an institutional repository for data sharing. National governments and the EU are key towards opening up data and alleviating third party constraints. They may also aid open data sharing by maintaining general repositories and offering facilitating solutions on a(n) (inter)national level.

6.7.2. Open access

Data analysis both reinforced and rejected claims by literature. Open access adoption by the respondent group proved to revolve around effort and publishing costs. Akin to open data, researchers exhibit a strong tendency towards avoiding additional effort to publish through open access. In relation to other attributes, effort surfaces as most important on a relative scale. The respondent group reacts negatively to increases in publishing costs. Although significantly less harmful than high levels of effort, publishing costs infuse reluctance to publish through open access. This research rejects the importance of visibility benefits, as negative utility is obtained from an increase in exposure. During experimentation, several respondents expressed they were unlikely to publish through open access, since they rated the available journals as low in quality. This notion may have led to underestimation of visibility gains for open access. Increases in academic recognition positively influence open access adoption as well, as where social engagement leaves adoption unbothered.

Disciplinary interaction effects yielded a rebuttal for claims made by Schöpfel et al. (2016) and Eger et al. (2016). As where interaction effects between technology and open access adoption were expected, their significance remained absent. Similar claims hold true for social science, although it was found academic recognition positively drives adoption within this discipline. The claim biologists are more receptive towards article publishing costs does not hold within the respondent group.

The policy implications are three-fold. First, research institutions, journals and governments alike should focus on lowering the bar for open access adoption. Research institutions may do so by installing supportive frameworks to educate and aid their staff on open access publishing. Journals may participate in doing so by entering agreements with research institutions, as well as offering support from their side as well. Governments may host (inter)national supportive agencies and NGOs for open access publishing as to spread awareness and forge alliances on a larger scale. Although the perceived quality of journals partially stems from longevity, journals and research institutions may attempt to strengthen their notoriety by obligatory open access publishing and expanding into research disciplines that currently do not offer a dominant open access journal. Grants and financers may increase adoption through stipulating open access adoption within their agreements. Thirdly, each stakeholder mentioned here is advised to contribute to lowering, financing or shifting the monetary burden associated with open access publishing. Researchers should not experience a financial barrier, as this negatively affects their likelihood to publish through open access journals.

7. Conclusion

This chapter contains final remarks on both the research process as well as its key findings. Here, the main findings are integrated into a set of concluding remarks and provides a crisp overview of the answer to the main research questions. Furthermore, the societal and theoretical relevance of this research are presented as well. Here, one may find how the main findings add to the current array of literature, as well as how they may affect society. Finally, this research is reflected upon, ahead of its limitations and opportunities for future research.

7.1. Conclusion & main findings

This research objective of this study was to gain an understanding of the dynamics underlying open science principle adoption by researchers. Before proceeding to the main findings of this research, it is first necessary to remove any ambiguity as to what open science entails in the light of this study. Since the behaviour of scholars is centralized, open science is narrowed down to the field of open research, which can be further divided into open access, open data and open source.

1. *What factors drive and inhibit the adoption of open science principles by researchers?*

During literature search, it became evident an abundance of factors considered influential towards open science adoption were widely described. However, the behavioural nature of the main research question required factors included for experimentation to be narrowed down to those related to decision-making. This approach established the need for a set of drivers and inhibitors to be considered for this research. For open data respectively, it was chosen to include 1) social engagement 2) effort 3) recognition 4) control and 5) data quality. With regard to open access, 1) social engagement 2) effort 3) visibility 4) recognition and 5) publishing costs were determined to comprise the factors that exemplify researcher behaviour accurately.

2. *What is the relative importance of factors that influence open science adoption?*

By means of stated choice experiments, it was shown that open science adoption is subject to external factors as much as the attributes specified above. Especially for openly sharing research data, it was found third party agreements and copyright concerns occupied a central role in decision-making, yet they were not (and could not) be included within the experiment. For open access, a variety of factors appeared relevant in guiding behaviour. The perceived effort to publish open access proved indicative of the likelihood to adopt open access publishing, as well as the costs associated with doing so. Contradicting expectations, an increase in visibility was determined to affect adoption negatively. Although remarkable, this enhanced the notion of external factors exemplifying behaviour for open access as well. In addition, it was observed respondents perceived subscription-based journals as more renowned, disclosing they would only publish in open access journals if no other options existed. Furthermore, it was concluded open access journals are not adequately represented within each field of study, rendering some respondents unable to adopt open access publishing.

3. *How does the relative importance of attributes vary across disciplines?*

A study of interaction effects between disciplines and open science principle adoption unveiled a multitude of insights. In terms of open access, disciplinary differences claimed throughout literature are mostly rejected. Positive interaction between technology and open access did not pertain to the results. Neither did clear positive correlation between social science and open access adoption stem from data analysis. In contrary, biologists proved more receptive to open access publishing on a relative scale. The claim biologists are more receptive to article publishing costs does not hold within the respondent group, rejecting claims by Schöpfel et al. (2016).

Discipline-based, quantitative research on open data adoption is sparse. Here, it is shown technology negatively interacts with open research data sharing. This may stem from the fact competition between research groups within this discipline is large. Social scientists exhibit a greater likelihood to engage in open data sharing, despite comments on third party data restrictions being prevalent within the respondent group.

4. *What are the policy implications of the attributes and their relative importance?*

The main findings result in a variety of policy implications. As where some recommendations are directed towards collaborative effort within the science community, others can be readily integrated into institutional policy. Academic recognition and financing have the potential to spearhead open science adoption amongst researchers. By devising a framework to finance open access publishing, researchers are less burdened with financial hardships. These remarks apply to both governing bodies such as national governments and the EU, financers and research institutions alike. Furthermore, recognizing an individual's effort to contribute to opening up science in the form of tenure and promotion, may pave the way for a future of freely sharing and publishing research as well. The role of research institutes and their corresponding policy should be that of facilitator. With proper support, infrastructure and academic rewards in place, institutions will act as an enabler rather than a hurdle towards open science adoption. It is however stressed that research institutions cannot accomplish the transition to an open academic community alone. Collaborative effort is required to alleviate barriers that prevent researchers from sharing data and improve the perceived quality difference between means of scholarly communication. It is therefore advised for institutions to establish partnerships with data agencies, financial institutions and form an integrative bond between the production and publication of novel scientific work. Financial institutions such as grants hold the regulatory power to stipulate open data sharing and open access publishing requirements within their agreements and are encouraged to do so. Open access journals face the issue of being perceived of lower quality than subscription-based journals. Disruptive agreements such as the Elsevier agreement change the dynamics of these issues and governments, along with national organizations are strongly advised to push for the expansion of such agreements. Journals themselves may seek to form agreements with research institutions on publishing costs and form frameworks to reduce effort. In terms of open data sharing, (inter)national governing bodies should prioritize maintaining repositories, alleviate data restrictions and enforce the regulatory power invested in them to expand current collaborative efforts further.

7.2. Societal relevance

Ultimately, the policy implications should serve as advisory tool for decision makers wanting to increase open science adoption, irrespective of the level of aggregations. The contributions of the preceding sub-questions yield potential learnings for governing bodies in the academic realm. Therefore, relevance stems from the transformation of outcomes into implications for governing bodies altogether. It is also noted that, in order to answer the main research question, it is strictly required to gain insights into the dynamics of open science adoption by researchers. In turn, the behavioural learnings produced by this study are estimated to propel forward the journey towards ubiquitously accessible scholarly communication. Crisply formulated: by investigating how those at the incipience of novel knowledge production can be understood behaviourally, this research contributes to the transformational movement of making knowledge available to everyone.

7.3. Theoretical relevance

In terms of contributions to literature, contributions are expected three-fold, representing novel knowledge added to each observed gap. First, findings are expected to provide an enhanced understanding of how the array of factors considered affect the behaviour of individual researchers in adopting open science principles. Rather than global trend observation, a behavioural, individualistic approach is added to current knowledge.

In terms of the knowledge gaps, several contributions are made. With regard to open access, the role of funding agencies and journals, exemplified by Kim & Adler (2015) is partially confirmed. In terms of quantification, the impact visibility benefits observed by Eysenbach (2006) and Creaser (2010) is sharply rejected. Respondents sharply negate the importance of visibility benefits in relationship to open access publishing. Claims by Björk (2004), Forrester (2015) and Klang et al. (2008) on the importance of (legal) support for open access publishing are reinforced and supported with quantitative data. Article processing costs are crucial towards the adoption of open access publishing, as stated by Nariani & Fernandez (2013) and Beasley (2016). Here, we show these claims hold and add it as a factor that ranks as less important than effort, yet more significant than any other factor subjected to research. Mercieca & Macauley (2008) states the perceived lower quality of open access journals exemplified throughout the respondent group. Therefore, this research reinforces these issues.

In terms of open data, the importance of culture of data sharing claimed by Sayogo & Pardo (2013) and Kim & Adler (2015) is not observed within the respondent group, thereby being insignificant. Findings by Kim & Adler (2015) and Campbell (2015) on the negative relationship between effort and open data sharing strongly persist within the respondent group. Hence, we prioritize 'Effort' as a factor to be incorporated into policy. Furthermore, remarks on data control and data quality by Fecher, Friesike & Hebing (2015) and Tenopir et al. (2011) are rejected based on the main findings, as data quality and control exhibited no significance to open data adoption. Despite it being observed as a general trend throughout literature, trust did not appear as a driver behind researcher behaviour.

Furthermore, by employing stated-choice experiments, this research will tread into the domain of quantitative research and allow for comparative examination of the drivers and inhibitors under scrutiny. Moreover, the methodology applied here is uncommon throughout literature as well. Finally, comparative data investigating the differences between research disciplines are added to contemporary literature as well. That is, rather than devising a prioritization unidimensionally, the effect of discipline-related researchers' background is to be considered as well.

On a final note, it is noted the current array of literature predominantly assumes a favourable stance with regard to open access publishing as point of departure. That is, rather than investigating whether researchers are likely to consider open access publishing, it directly delves into benefits and trend observation. During experimentation, it was found a group of respondents would only publish in open access journals, if no other options were available. As where one would expect current research to be framed according to such notions, this stage is largely skipped. Contributions by Schöpfel et al. (2016) and Eger et al. (2016) find itself at the incipience of distinguishing such patterns, yet fail to relate their findings accordingly. Hence, this research suggests that future work retreats to the stage of how open access may surpass subscription-based publishing, rather than assuming it as the de facto standard.

7.4. Limitations

Both the research design and methodology are subject to a variety of limitations. The most apparent limitation is the timeframe available for this research. As can be evidently seen from the discrepancy between the number of desired responses and the number of actual responses, the window for data collection was rather slim. Although this ranges back to a narrow research demarcation, limited time availability is rendered as a significant impact factor to the research process.

In terms of methodology, a few limitations apply. Stated-choice experiments excel at unveiling the relative importance of attributes included for experimentation. However, it abstains from reporting absolute utility or likelihood. That is, as where the significance of factors could be readily assessed, we cannot draw conclusions on absolute adoption size. Therefore, one may not state adoption within discipline group A is larger than discipline group B, for example. Although the methodology fits the (sub) research questions, this is considered a pitfall. Furthermore, tediousness plays a significant role in the number of attributes that may be included for experimentation. Too many attributes will lead to respondent exhaustion, inducing unrealistic behaviour. Therefore, only a selection of attributes could be included for experimentation, rather than the full set of attributes available. Besides, attributes have to reflect decision variables that may be influence by either behaviour or policy. This disregarded the inclusion of factors such as the availability of an open access journal to publish in or freedom of third party data restrictions. Moreover, the personal background of researchers was not incorporated into experimentation. Therefore, no conclusions can be drawn on how different personal values, background and beliefs affect adoption.

Since this research is directed towards Dutch universities and respondents affiliated with institutions from this country, main findings should be interpreted as such. Academic policy enforced by Dutch universities as well as national policy is likely to have affected the main findings. Therefore, one should not extrapolate conclusions drawn here and apply them in a different setting without additional scrutiny. Both the facilitating conditions available and regulatory frameworks are likely to differ in different geographical regions. Furthermore, Schöpfel et al. (2016) found discrepancies induced by cultural differences between France and the USA over the course of his research. This exemplifies cultural values are not negligible with regard to open science principle adoption. Applicability of the main findings is not claimed to limit to the Netherland only, yet one should tread with caution if applying our results elsewhere.

7.5. Connection to MSc Engineering & Policy Analysis

This study reflects a manifold of concepts introduced over the course of the Engineering & Policy Analysis master program. Regardless of a sub-research question dedicated to the policy implications of the main findings, the nature of the problem under scrutiny stems from a multi-actor context. Not only is the behaviour of individual researchers measured against that of individual bodies such as research institutions, financial institutions, scientific journals and data repositories, their underlying beliefs, values and actions do not necessarily align. Ranging back to the attributes included in stated choice experiments, 'Effort' – the support offered by research institutions and 'Control' – the attitude of data repositories towards data submitted through them, exhibit a tendency towards different interests. Although the methodology applied here is rather alien to the Engineering & Policy Analysis program itself, one could argue it incorporates the desired analytical character. Despite the fact choice models and their simulation only operate on the interface between the modelling taught during EPA courses, the subject under study required the development and adaptation of computer models in a certain regard.

Open science does not fit into the traditional definition of grand challenges, as one could argue multiple rounds of action exist and action taken is not irrevocable. Nonetheless, the advancement of science into a new era of openness is a contemporary issue, which is arguably necessary for progressing society further. Moreover, open science allows for enhanced collaboration and transparency, which could in turn be applied to alleviate traditional grand challenges. That is to say, open science does not lead the list of challenges posed to us by modern society, yet it fits the rationale utilized to label other challenges as grand.

7.6. Reflection

With the hypernym that is open science, comes compromise in terms of research scope. One seeks to obtain actionable, insightful results, whilst avoiding myopia on a certain aspect of open science. Here, it was opted for to explore open science adoption emphasizing on open data and open access. Although this distinction is thoroughly grounded in literature, one may argue that open science adoption by researchers transcends the realm of open data and open access alone, extending into, amongst others, open source as well. Although the importance of open source as impetus to open data is undeniably recognized throughout this research, other attributes were prioritized for experimentation. During literature search, it became apparent behavioural determinants for open data adoption were heavily nested within technology and university policy & rewards. Moreover, the connection between open data and open access in the light of open science could be readily established. This study and its main research question revolve around a behavioural examination of the drivers and inhibitors of open science adoption, less so than investigating the relative importance of individual components of open research itself. Though such a distinction impedes knowledge gained on certain aspects of open research, it manages to do so for the fields of open access and open data.

The methodology of choice poses multiple implications to research outcomes. Stated choice experiments allow for the inclusion of a certain set of attributes only. That is, the amount of attributes as well as the type of attributes that can be included are limited. With the inclusion of factors responsive to policy or behaviour, an array of factors observed relevant by literature are rendered obsolete. Hence, more drivers & inhibitors than included for experimentation affect researchers' behaviour, yet they cannot be comparatively examined through stated choice experiments. A prime example is the perceived quality of open access journals. Although quality-related attributes were included, a true comparison between subscription-based journals and open access journals was left unconsidered for experimentation. Therefore, the quality of both journals is assumed equal throughout the experiment. In reality, this claim does not hold indefinitely. Moreover, multiple respondents have indicated not to publish through open access due to perceived quality differences. It is therefore likely that studies employing other research methods yield different results, if they were to evaluate such disparate factors. Nonetheless, researcher preferences obtained through stated choice experiments under the attributes of choice are highly actionable for policy formation and exemplify decision-making in adopting open science principles.

On a more personal note, this research also requires reflection in terms of process. One could say several hardships were overcome ahead of writing this very reflection. As where it was already known the agenda was rather tight, it was proven more difficult to complete certain phases than expected. Most impactful was the sluggish speed at which responses were amassed. Despite exhaustive effort to share the experiment, the number of respondents lagged behind. Due to time constraints, it was decided to terminate data collection as to proceed to data analysis. With the incurred delay, data analysis therefore became more restricted in terms of time available. Although it is not reflected upon as unsatisfactory, a more balanced planning would be more appropriate.

Nonetheless, the hardships experienced during data collection are attributed to the research demarcation. As where crisp, well-defined boundaries can serve to enhance clarity, they may also become limiting in another sense. Addressing both cultural differences and disciplinary differences has proven too ambitious to be carried out within the time available for this research. The amount of respondents that could be gathered by limiting the research population to the Netherlands was highly overestimated. The narrowly defined research demarcation is considered the main flaw of this research, impeding the analysis of discipline-specific behavioural patterns. Despite there being clear lessons for future projects, the positive experience of being able to execute, manage and document this research will be cherished by the author.

7.7. Opportunities for future research

This research yields a first exploration of the behavioural domain of open science principle adoption by researchers. As where a solid knowledge base has been established by means of this study, it offers a plethora of opportunities for future research. The narrow focus on open data and open access as classification for open research – and thereby open science, foregoes the importance of other aspects of open science (and research). Future research may be directed towards the investigation of the behavioural implications of open source adoption with regard to open research as a whole. Contributions within the field of the open data-open source connection are largely unprecedented, especially in the behavioural sense.

As noted before, only general attributes were included for experimentation under a limited number of attribute levels. One may seek to delve deeper into the attributes found most influential during this study. In addition, the behavioural, preference-based approach attained here is yielded to act as a stimulus for transformative action within the field of open science adoption research. As where the current array of literature is predominantly of descriptive nature, actionable, policy-fuelling research will most likely spawn from the domain of preference-based methods. That is to say, in order for research to contribute to its own openness, future work is to be directed towards unveiling the constraints of adopting a paradigm that has amassed almost unanimous support from a group in which adoption itself is lagging behind.

Appendix I: Survey Questions

The stated-choice experiment consists of multiple parts and will gather comparative information on which factors researchers deem more important than others. Preceding those questions, participants are prompted for demographic information. The stated choice experiments are to be further divided into three parts: open data, open access and questions on grants & institutions.

Demographics & Background

1. Are you currently affiliated with a Dutch university?

   Yes -> continue

   No -> You do not belong to the target group of this research

2. What is your age?

   Division into the following age groups: 18-29, 30-39, 40-49, 50-59, 59-65, 65+

3. In which country do you reside?

   [Drop down list of countries]

4. What is the highest level of education you have completed or the highest degree you have received?
   - Elementary school
   - High school (or equivalent)
   - Associate's degree
   - Bachelor's degree
   - Master's degree
   - PhD
   - Other (please specify)

5. What is your current position?
   - PhD candidate
   - Post-doctoral researcher
   - Assistant professor
   - Associate professor
   - Full professor
   - Other

6. Which research discipline do you consider yourself to belong to?

   Choice from the list provided by the NWO, included as a list in Appendix I.I.

   Other, namely….

7. Which university are you currently affiliated with

   Open question.

Experience

Open data

8. Have you ever openly shared one or more of your research datasets through a repository or in any other way? (e.g. through the repository of the 4TU.Centre for Research Data, DANS, on your own website, CORE or ArXiv)

   a. Yes, I have shared one or more of my datasets openly (through either repositories or other ways)
   b. No, I have never shared any of my datasets freely
   c. I am unsure whether any of my self-collected datasets were published openly

9. If yes, how would you rank your involvement with *sharing* open research data?

   Score from 1-10, 1 being 'barely' and 10 being 'common user'

10. Have you ever used (e.g. viewed, browsed, analyzed, visualized, compared, enhanced) a research dataset that someone else had openly shared (e.g. through a research data repository)?

    a. Yes, I have used datasets others have openly shared to the benefit of my own research
    b. No, I have not consulted any openly shared datasets [Go to Question 7]
    c. I am unsure whether I have ever incorporated open datasets in my research.

11. If yes, how would you rank your involvement with *using* open data?

    Score from 1-10, 1 being 'barely' and 10 being 'common user'

Open access

12. Which proportion of your publications did you, on average, publish using open access?

    a. 0-20%
    b. 21-40%
    c. 41-60%
    d. 61-80%
    e. 81-100%

13. Have you ever used (e.g. read, viewed, referenced, compared, peer-reviewed) an open access publication?

    a. Yes, I have used open access publications
    b. No, I have not used any openly published work [Go to Question 10]
    c. I am unaware of whether I ever used open access publications

14. How would you rank your involvement with *using* open access publications? (Note: I feel as if Q12 incorporates the degree of involvement for publishing already, hence I decided to not ask this question for open access publishing).

    Score from 1-10, 1 being 'barely' and 10 being 'common user'

Grants & Institutions

15. If a grant would require you to openly share your research data. How likely would you be to obey this requirement and publish all of your datasets openly?

    Score from 1-10, 1 being 'highly unlikely' and 10 being 'most likely'

16. If a grant would require you to publish your research financed by that grant through Open Access. How likely would you be to obey this requirement and publish all of your work through Open Access?

    Score from 1-10, 1 being 'highly unlikely' and 10 being 'most likely'

Appendix I.I. List of NWO disciplines

- Archaeology
- Area studies
- Art and architecture
- Astronomy, astrophysics
- Biology
- Business administration
- Chemistry
- Communication science
- Computer science
- Computers and the humanities
- Cultural anthropology
- Demography
- Development studies
- Earth sciences
- Economics
- Educational Sciences
- Environmental science
- Gender studies
- Geography / planning
- History
- History of science
- Language and literature
- Law
- Life sciences
- Life sciences and medicine

- Linguistics
- Mathematics
- Medicine
- Music, theatre, performing arts and media
- Pedagogics
- Philosophy
- Physics
- Psychology
- Public administration and political science
- Religious studies and theology
- Science and technology
- Sociology
- Technology
- Veterinary medicine

Appendix II: Stated-Choice questions

This appendix holds the full description of the stated-choice experiment choice tasks, the syntax used in Ngene as well as the corresponding designs and their efficiency measures.

Syntax, Design & Efficiency measures in Ngene

Figure II.I and figure II.II hold the syntax utilized for the generation of the final experiment designs.

```
Design
;alts = alt1, alt2
;rows = 10
;eff = (mnl,d)
;model:
U(alt1) = b1[0.01]*A[1,2,3]+b2[-0.01]*B[1,2,3]+b3[0.01]*C[1,2,3]+b4[0.01]*D[1,2,3]+b5[-0.01]*E[1,2,3]/
U(alt2) = 0
$
```

Figure II.I Open Access experiment design syntax

```
Design
;alts = alt1, alt2
;rows = 10
;eff = (mnl,d)
;model:
U(alt1) = b1[0.01]*A[1,2,3]+b2[-0.01]*B[1,2,3]+b3[0.01]*C[1,2,3]+b4[0.01]*D[1,2,3]+b5[0.01]*E[1,2,3]/
U(alt2) = 0
$
```

Figure II.II Open Data experiment design syntax

MNL efficiency measures					
D error	0.58245				
A error	0.584665				
B estimate	99.909824				
S estimate	22686.302998				
Prior	b1	b2	b3	b4	b5
Fixed prior value	0.01	-0.01	0.01	0.01	0.01
Sp estimates	22686.302998	22308.61314	22310.195443	22311.152236	22686.205535
Sp t-ratios	0.013013	0.013123	0.013122	0.013122	0.013013
Design					
Choice situation	alt1.a	alt1.b	alt1.c	alt1.d	alt1.e
1	3	3	2	3	1
2	2	3	3	1	3
3	1	3	2	1	1
4	3	1	2	1	3
5	2	1	1	1	1
6	3	2	1	2	2
7	1	1	3	2	1
8	1	2	1	2	2
9	1	2	1	3	3
10	2	1	3	3	2

Figure II.III Open data choice tasks & efficiency measures

MNL efficiency measures					
D error	0.582015				
A error	0.584228				
B estimate	99.981755				
S estimate	22670.752158				
Prior	b1	b2	b3	b4	b5
Fixed prior value	0.01	-0.01	0.01	0.01	-0.01
Sp estimates	22670.357087	22292.239035	22293.026399	22670.752158	22292.233256
Sp t-ratios	0.013017	0.013127	0.013127	0.013017	0.013127
Design					
Choice situation	alt1.a	alt1.b	alt1.c	alt1.d	alt1.e
1	3	1	2	2	3
2	1	2	1	1	1
3	2	2	3	1	2
4	1	3	1	3	3
5	2	3	3	2	1
6	3	3	1	1	2
7	1	1	2	1	3
8	3	1	1	3	1
9	1	1	2	2	1
10	2	2	3	3	2

Figure II.IV Open access choice tasks & efficiency measures

Choice Task Tables

The mathematical design obtained through Ngene maps the levels and attributes onto several choice a set of choice tasks. Although figure II.III & II.IV include those mappings, Table II.I & II.II holds a full overview of the choice tasks, with the attributes labelled correctly. Each row specifies a choice task under the levels stated for each corresponding attribute. Table II.I represents choice tasks for open data, as where Table II.II concerns choice tasks for open access.

	Social Engagement	Effort	Recognition	Control	Data Quality
Choice task 1	3	3	2	3	1
Choice task 2	2	3	3	1	3
Choice task 3	1	3	2	1	1
Choice task 4	3	1	2	1	3
Choice task 5	2	1	1	1	1
Choice task 6	3	2	1	2	2
Choice task 7	1	1	3	2	3
Choice task 8	1	2	1	2	2
Choice task 9	1	2	1	3	3
Choice task 10	2	1	3	3	3

Table II.I Open data choice tasks

	Social Engagement	Effort	Visibility	Recognition	Publishing costs
Choice task 1	3	1	2	2	3
Choice task 2	1	2	1	1	1
Choice task 3	2	2	3	1	2
Choice task 4	1	3	1	3	3
Choice task 5	2	3	3	2	1
Choice task 6	3	3	1	1	2
Choice task 7	1	1	2	1	3
Choice task 8	3	1	1	3	1
Choice task 9	1	1	2	2	1
Choice task 10	2	2	3	3	2

Table II.II Open access choice tasks

Appendix III: Source code data manipulation & MNL model

This section holds the source code used to i) manipulate the data set ii) run MNL models for both open data and open access.

```
library(ggplot2)

library(lattice)

library("readxl")

library("writexl")

library("plyr")

library(psych)

library("mltools")

library(data.table)

library(operators)


# Reading in Data from file in both numbers and with answer text

mydata_utf8_statedchoice =
read.table("/Users/maart/Documents/R/Data/ResultsJuly11.csv",sep=",",fileEncoding = "UTF-8-
BOM",header=TRUE)

stated_choice_matrix = read_xlsx("/Users/maart/Documents/R/Data/StatedChoiceMatrix.xlsx")


#set cols for discipline-based filters

#social science

dem_col = c(37,36,27,40)


#technology (incl. computer science/mathematics)

dem_col = c(10,28,35,39,41)


#biology

dem_col = c(26,25,29)


#economics

dem_col = c(7,16)


# drop off metadata columns

mydata_statedchoice = mydata_utf8_statedchoice

names(mydata_statedchoice)[names(mydata_statedchoice) == "X6_1"] <- "Discipline"
```

Source code 1 Data manipulation

```
### Load Apollo library
library(apollo)
library(data.table)



### Initialise code
apollo_initialise()

### Set core controls
apollo_control = list(
  modelName  ="MNL_OA",
  modelDescr ="MNL model OA",
  indivID    ="ID"
)

#### LOAD DATA, read in onehot from file
database = onehot_OA



### Vector of parameters, including any that are kept fixed in estimation
apollo_beta=c(BETA_Att12   = 0,
        BETA_Att13   = 0,
        BETA_Att22   = 0,
        BETA_Att23   = 0,
        BETA_Att32   = 0,
        BETA_Att33   = 0,
        BETA_Att42   = 0,
        BETA_Att43   = 0,
        BETA_Att52   = 0,
        BETA_Att53   = 0,
        BETA_Disc    = 0,
```

Source code 2 MNL model source code:l open access

```
### Load Apollo library
library(apollo)

### Initialise code
apollo_initialise()

### Set core controls
apollo_control = list(
  modelName  ="MNL_OD_SS",
  modelDescr ="MNL model OD",
  indivID    ="ID"
)

#### LOAD DATA, read in onehot from file
database = onehot_OD



### Vector of parameters, including any that are kept fixed in estimation
apollo_beta=c(BETA_Att12   = 0,
        BETA_Att13   = 0,
        BETA_Att22   = 0,
        BETA_Att23   = 0,
        BETA_Att32   = 0,
        BETA_Att33   = 0,
        BETA_Att42   = 0,
        BETA_Att43   = 0,
        BETA_Att5    = 0,
        BETA_Disc    = 0,
        ASC_A = 0)
```

Source code 3 MNL Model: open data

Appendix IV: Literature Review Search Terms

In order to compose a comprehensive literature of contemporary sources, a plethora of search terms was utilized. This section enumerates the search terms used during this process.

Open science	"Open science value" "Open science definition" "Defining open science" "Open science" AND "brands" "Open science" AND "barriers" "Barriers to open research" "Incentivizing open science"
Open access	"Open access" "Open access" AND "drivers" "Open access advantages" "Open access adoption" "Open access" AND "adoption" "Open access" AND "researcher preferences" "Open access" AND "visibility" "Open access" AND "incentives" "Open access" AND "academic recognition" "Open access" AND "business models" "Article processing costs" AND "Open access" "Open access drivers" "Open access adoption discipline differences"
Open data	"Open data" AND "drivers" "Open data adoption" "Open data" AND "researcher preferences" "Open data" AND "control" "Open data" AND "integrity" "Open data" AND "discipline"
Open source	"Open source" AND "open science" "Open source" AND "reproducibility" "Executable papers" "Open source" AND "advantages" "Open source benefits" "Open source advantages" "Open source" AND "transparency" "Open source collaboration"

References

Adamowicz, W., Louviere, J., & Swait, J. (1998). Introduction to attribute-based stated choice methods. *A report submitted to: National Oceanic and Atmospheric Administration (NOAA)*. Ch.3.

Ale Ebrahim, N., Salehi, H., Embi, M. A., Habibi, F., Gholizadeh, H., & Motahar, S. M. (2014). Visibility and citation impact. International Education Studies, 7(4), 120-125.

Ali-Khan, S.E., Harris, L.W., & Gold, E.R. (2017). Motivating participation in open science by examining researcher incentives. *eLife*.

Allen, J. (2005). Interdisciplinary differences in attitudes towards deposit in institutional repositories (Doctoral dissertation, Manchester Metropolitan University (UK)).

Arza V, Fressoli M. (2017) Systematizing benefits of open science practices. Information Services & Use.;37(4):463-74

Arzberger P, Schroeder P, Beaulieu A, Bowker G, Casey K, Laaksonen L, et al. Promoting access to public research data for scientific, economic, and social development. Data Science Journal. 2004;3(29):135-52.

Beasley, G. (2016). Article processing charges: a new route to open access? *Information Services and Use*, 36(3-4), 163–170. https://doi.org/10.3233/ISU-160815

Beecher, K., Boldyreff, C., Capiluppi, A., & Rank, S. (2008). Evolutionary success of open source software: An investigation into exogenous drivers. *Electronic Communications of the EASST*, *8*.

Björk, B.-C. Open access to scientific publications: An analysis of the barriers to change. Inf. Res.: Int. Electron. J. 2004, 9, 170.

Björk, B. (2013). Open Access - Are the Barriers to Change Receding? *Publications, 1*, 5-15.

Bliemer, M. C., & Rose, J. M. (2005). Efficiency and sample size requirements for stated choice studies.

Borgman, C. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology, 63*(6), 1059-1078. doi:10.1002/asi.22634

Borgman, C. (2007) "The View From Here." In Scholarship in the Digital Age: Information, Infrastructure, and the Interne, 227–261. Cambridge, MA: MIT Press.

Bornmann, L., & Daniel, H. D. (2007). What do we know about the h index?. *Journal of the American Society for Information Science and technology*, 58(9), 1381-1385.

Campbell J. (2015) Access to scientific data in the 21st century: Rationale and illustrative usage rights review. *Data Science Journal*;13,203–30.

Ceci, S. J. (1988). Scientists' attitudes toward data sharing. Science, Technology, & Human Values, 13(1/2), 45-52

Chorus, C. (2020) Lecture on Discrete Choice Modelling [Powerpoint slides]

Creaser, C. (2010). Open access to research outputs-institutional policies and researchers' views: Results from two complementary surveys. *New Review of Academic Librarianship, 16*(1), 4-25. doi:10.1080/13614530903162854

Curty, R., Crowston, K., Specht, A., Grant, B., Dalton, E., & Sugimoto, C. (2017). Attitudes and norms affecting scientists' data reuse. *Plos One, 12*(12), 0189288. doi:10.1371/journal.pone.0189288

David, P. A., den Besten, M.D., & Schroeder, R. (2009). Collaborative research in eScience and open access to information. SIEPR Discussion Paper No. 08-21. Stanford University.

Davis, P.M., & Connolly, M.J. (2007). Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace. *D-Lib Mag., 13*.

De Jong, M. (2019) Lectures notes on Open Science: Sharing Your Research with the World. Retrieved March 30, 2020, from https://online-learning.tudelft.nl/courses/open-science-sharing-your-research-with-the-world/

Doyle, H., Gass, A., & Kennison, R. (2004). Who pays for open access?. *PLoS Biology*, *2*(4).

Dr Hwang and the stem cell swindle. (2011, October 23). Retrieved March 23, 2020, from https://www.independent.co.uk/news/science/dr-hwang-and-the-stem-cell-swindle-1809943.htm

Dulle, F. W., Minish-Majanja, M. K., & Cloete, L. M. (2010). Factors influencing the adoption of open access scholarly communication in Tanzanian public universities. In *World library and information congress: 76th IFLA general conference and assembly* (pp. 10-15).

Easterbrook, S.M. (2014). Open code for open science? *Nature Geoscience*, *7*, 779–781. doi:10.1038/ngeo2283. .

Enke, N., Thessen, A., Bach, K., Bendix, J., Seeger, B., & Gemeinholzer, B. (2012). The user's view on biodiversity data sharing — investigating facts of acceptance and requirements to realize a sustainable use of research data —. *Ecological Informatics*, 11, 25-33. doi:10.1016/j.ecoinf.2012.03.004

Eger, T., Scheufen, M., & Meierrieks, D. (2016). The determinants of open access publishing: Survey evidence from countries in the mediterranean open access network (medoanet). *Economia E Politica Industriale : Journal of Industrial and Business Economics, 43*(4), 463-489. doi:10.1007/s40812-016-0027-8

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS biology*, 4(5).

Fecher, B., & Friesike, S. (2013). Open Science: One Term, Five Schools of Thought.

Forrester, A. (2015). Barriers to Open Access Publishing: Views from the Library Literature. *Publications, 3*, 190-210.

Franzen, M., Rödder, S., & Weingart, P. (2007). Fraud: Causes and culprits as perceived by science and the media: Institutional changes, rather than individual motivations, encourage misconduct. *EMBO reports*, *8*(1), 3-7.

Fraude hoogleraar Stapel 'verbijsterend'. (n.d.). Retrieved March 23, 2020, from https://nos.nl/artikel/308864-fraude-hoogleraar-stapel-verbijsterend.htm

Fry, J., Oppenheim, C., Probets, S., Creaser, C., Greenwood, H., Spezi, V. and White, S. (2010), *PEER behavioural research: Authors and users vis-à-vis journals and repositories. Baseline report*. LISU Loughborough University, available at http://www.peerproject.eu/fileadmin/media/reports/Final_revision_-_behavioural_baseline_report_-_20_01_10.pdf

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. PloS one, 5(10), e13636.

Getz, M. (2005). Open-access scholarly publishing in economic perspective. *Journal of Library Administration*, *42*(1), 1–39. https://doi.org/10.1300/J111v42n01_01

Gezelter, J. D. (2015). Open source and open data should be standard practices.

Guédon Jean-Claude. (2004). The “green” and “gold” roads to open access: the case for mixing and matching. Serials Review, 30(4), 315–328. https://doi.org/10.1080/00987913.2004.10764931

Haider, W. (2002). Stated preference and choice models–A versatile alternative to traditional recreation research. In *Monitoring and management of visitor flows in recreational and protected areas. Conference proceedings* (pp. 115-121).

Harper, L., & Kim, Y. (2018). Attitudinal, normative, and resource factors affecting psychologists' intentions to adopt an open data badge: An empirical analysis. *International Journal of Information Management*, 41, 23-23.

Haslinger, I.M. (2019) TU Delft Strategic Plan Open Science 2020-2024: Research and Education in the Open Era. Retrieved from https://repository.tudelft.nl/islandora/object/uuid:f2faff07-408f-4cec-bd87-0919c9e4c26f?collection=research (August 2020).

Hauber, A. B., González, J. M., Groothuis-Oudshoorn, C. G., Prior, T., Marshall, D. A., Cunningham, C., IJzerman, M. J., & Bridges, J. F. (2016). Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, *19*(4), 300–315. https://doi.org/10.1016/j.jval.2016.04.004

Henriksen-Bulmer, J., & Jeary, S. (2016). Re-identification attacks - a systematic literature review. *International Journal of Information Management, 36*(6), 1184-1184. doi:10.1016/j.ijinfomgt.2016.08.002

Hey, T., & Payne, M. C. (2015). Open science decoded. *Nature physics*, *11*(5), 367-369.

Hess, S. & Palma, D. (2019). Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application, *Journal of Choice Modelling*, *32*, 100170

Hofstede, G. H., Hofstede, G. J., & Minkov, M. (2010). Cultures and organizations: Software of the mind. *Maidenhead: McGraw-Hill*, 341-371.

Horta, H., & Santos João M. (2016). The impact of publishing during phd studies on career research publication, visibility, and collaborations. *Research in Higher Education*, 57(1), 28–50.

Huang X, Hawkins BA, Lei F, Miller GL, Favret C, et al. (2012) Willing or unwilling to share primary biodiversity data: results and implications of an international survey. *Conserv Lett 5*: 399–406. doi: 10. 1111/j.1755-263X.2012.00259.x

Jillian, C., Elizabeth, R., & Christine, L. (2013). If we share data, will anyone use them? data sharing and reuse in the long tail of science and technology. *Plos One, 8*(7). doi:10.1371/journal.pone.0067332

Joo S, Kim S, Kim Y. (2017) An exploratory study of health scientists' data reuse behaviors: Examining attitudinal, social, and resource factors. *Aslib Journal of Information Management*. ;69(4):389-407.

Kauppinen, T., & Espindola, G. M. de. (2011). Linked open science-communicating, sharing and evaluating data, methods and results for executable papers. *Procedia Computer Science*, *4*, 726–731. https://doi.org/10.1016/j.procs.2011.04.076

The key areas. (n.d.). Retrieved from https://www.openscience.nl/en/national-platform-open-science/the-key-areas

Kim Y, Adler M. (2015) Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *International Journal of Information Management*.;35(4):408-18

Klang, M., & Avgerou, Chrisanthi, London School of Economics and Political Science, UK. (2008). Social dimensions of information and communication technology policy : proceedings of the eighth international conference on human choice and computers (hcc8), ifip tc 9, pretoria, south africa, september 25-26, 2008. In *Open access barriers: an action research* (pp. 335–348). essay, Boston, MA : Springer US. https://doi.org/10.1007/978-0-387-84822-8_23

Kløjgaard, M.E., Bech, M., & Søgaard, R. (2012). Designing a stated choice experiment: The value of a qualitative process.

Koop, D., Santos, E., Mates, P., Vo, H. T., Bonnet, P., Bauer, B., … Silva Cláudio T. (2011). A provenance-based infrastructure to support the life cycle of executable papers. *Procedia Computer Science*, *4*, 648–657. https://doi.org/10.1016/j.procs.2011.04.068

Kraus, J. (2014). Cash, carrots, and sticks: open access incentives for researchers. *F1000research*, *3*, 263–263. https://doi.org/10.12688/f1000research.5242.1

Landry, R., Traore, N., & Godin, B. (1996). An econometric analysis of the effect of collaboration on academic research productivity. *Higher education*, *32*(3), 283-301

Leonelli, S., Spichtinger, D., & Prainsack, B. (2015). Sticks and carrots: Encouraging open science at its source. *Geo : Geography and Environment, 2*(1), 12-16.

Lindman, J. (2014). Similarities of open data and open source: Impacts on business. *Journal of theoretical and applied electronic commerce research*, *9*(3), 46-70.

Lindman, J., & Nyman, L. (2014). The businesses of open data and open source: Some key similarities and differences. *Technology Innovation Management Review*, *4*(1).

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: International Journal of an Emerging Transdiscipline*, 9(1), 181-212

Levin, N., Leonelli, S., Weckowska, D., Castle, D., & Dupré, J. (2016). How Do Scientists Define Openness? Exploring the Relationship Between Open Science Policies and Research Practice. *Bulletin of Science, Technology & Society*, 36(2), 128-141. doi:10.1177/0270467616668760

Lippert, J., Burghaus, R., Edginton, A., Frechen, S., Karlsson, M., Kovar, A., Lehr, T., Milligan, P., Nock, V., Ramusovic, S., Riggs, M., Schaller, S., Schlender, J., Schmidt, S., Sevestre, M., Sjögren, E., Solodenko, J., Staab, A. and Teutonico, D. (2019), Open Systems Pharmacology Community—An Open Access, Open Source, Open Science Approach to Modeling and Simulation in Pharmaceutical Sciences. CPT Pharmacometrics Syst. Pharmacol., 8: 878-882. doi:10.1002/psp4.12473

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: analysis and applications*. Cambridge university press. Ch.2.

Lwoga, E. T., & Questier, F. (2014). Faculty adoption and usage behaviour of open access scholarly communication in health science universities. *New Library World,* 115(3/4), 116–139. https://doi.org/10.1108/NLW-01-2014-0006

Lyon, L. (2016). Transparency: the emerging third dimension of Open Science and Open Data. *Liber quarterly*, *25*(4).

McCormick, M., Liu, X., Jomier, J., Marion, C., & Ibanez, L. (2014). ITK: enabling reproducible research and open science. *Frontiers in Neuroinformatics*, *8*. https://doi.org/10.3389/fninf.2014.00013

McGinn, C. (2012). *Basic structures of reality: Essays in meta-physics*. New York: Oxford University Press. Ch. 8

Mercieca, P., & Macauley, P. (2008). A New Era of Open Access?. *Australian Academic & Research Libraries*, 39(4), 243-252

Moed, H. F. (2007). The effect of "open access" on citation impact: an analysis of arxiv's condensed matter section. *Journal of the American Society for Information Science and Technology*, 58(13), 2047–2047.

Molin, E. (2017) Lecture on Efficient Design [Powerpoint slides]

Mosconi, G., Li, Q., Randall, D., Karasti, H., Tolmie, P., Barutzky, J., ... & Pipek, V. (2019). Three gaps in opening science. *Computer Supported Cooperative Work (CSCW)*, *28*(3-4), 749-789.

Migheli, M., & Ramello, G. B. (2014). Open access journals and academics' behavior. *Economic Inquiry*, 52(4), 1250–1250.

Mooney H, Newton MP. The anatomy of a data citation: *Discovery, reuse, and credit. Journal of Librarianship and Scholarly Communication*. 2012;1(1):eP1035

Murray-Rust, P. (2008). Open data in science. *Serials Review*, 34(1), 52-64. doi:10.1016/j.serrev.2008.01.001

Mosconi, G., Li, Q., Randall, D., Karasti, H., Tolmie, P., Barutzky, J., . . . Pipek, V. (2019). Three gaps in opening science. *Computer Supported Cooperative Work (cscw) : The Journal of Collaborative Computing and Work Practices, 28*(3-4), 749-789. doi:10.1007/s10606-019-09354-z

Narayan, B., & Luca, E. (2017). Issues and challenges in researchers' adoption of open access and institutional repositories: a contextual study of a university repository.

Nariani, R., & Fernandez, L. (2012). Open access publishing: what authors want. *College & Research Libraries*, 73(2), 182–195.

Open Science. (n.d.). Retrieved from https://ec.europa.eu/research/openscience/index.cfm

Open Access Figures in the Netherlands (n.d.). Retrieved from https://www.openaccess.nl/en/events/open-access-figures-in-the-netherlands

Open Access - Leiden University. (n.d.). Retrieved August 05, 2020, from https://www.library.universiteitleiden.nl/researchers/open-access

Open Access publishing. (n.d.). Retrieved August 05, 2020, from https://www.tudelft.nl/en/library/current-topics/libraryfor-researchers/library-for-researchers/publishing-outreach/open-access-publishing/

Orme, B. (2019) Sample Size Issues for Conjoint Analysis Studies, *Sawtooth Software Technical Paper*, Sequim. Retrieved from https://sawtoothsoftware.com/resources/technical-papers/sample-size-issues-for-conjoint-analysis-studies-2019

Piwowar, H. A., & Chapman, W. W. (2008). A review of journal policies for sharing research data. *Paper presented at the International Conference on Electronic Publishing*

Pinfield, S. (2010). Paying for open access? Institutional funding streams and OA publication charges. *Learned Publishing*, 23(1), 39-52.

Pinfield, S. (2015). Making open access work: the "state-of-the-art" in providing open access to scholarly literature. *Online Information Review, 39(*5), 604–636. https://doi.org/10.1108/OIR-05-2015-0167

Price-Whelan, A. M., Sipőcz B. M, Günther H. M, Lim, P. L., Crawford, S. M., Conseil, S., … Zabalza, V. (2018). The astropy project: building an open-science project and status of the v2.0 core package. *The Astronomical Journal*, *156*(3), 123–123. https://doi.org/10.3847/1538-3881/aabc4f

Publish Open Access - the agreements with publishers: Open Access discount. (n.d.). Retrieved August 05, 2020, from https://libguides.rug.nl/umcg/openaccess

Publishing in Open Access. (n.d.). Retrieved August 05, 2020, from https://www.eur.nl/en/research/research-services/publishing/open-access/publishing-open-access

Research Data Policy. (2015). Retrieved August 05, 2020 from https://www.rug.nl/research/research-data-management/downloads/c3-policy-dl/2015-research-data-policy.pdf

Rose, J. M., & Bliemer, M. C. (2013). Sample size requirements for stated choice experiments. *Transportation, 40*(5), 1021-1041. doi:10.1007/s11116-013-9451-z

Rose, J. M., & Bliemer, M. C. J. (2013). Sample size requirements for stated choice experiments. *Transportation*, *40*(5), 1021–1041. https://doi.org/10.1007/s11116-013-9451-z

Sarewitz, D. (2016). Saving science. *The New Atlantis*, 49(49), 4-40.

Sayogo, D., & Pardo, T. (2013). Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly, 30*, 31. doi:10.1016/j.giq.2012.06.011

Schöpfel, J., Ferrant, C., André, F., & Fabre, R. (2016). Ready for the Future? A Survey on Open Access with Scientists from the French National Research Center (CNRS).

Schoonen, W. (2020). Universiteiten en Elsevier sluiten 'unieke' deal over open access. 'Nederlandse wetenschap is voorbeeld in de wereld'. Retrieved August 06, 2020, from https://www.trouw.nl/nieuws/universiteiten-en-elsevier-sluiten-unieke-deal-over-open-access-nederlandse-wetenschap-is-voorbeeld-in-de-wereld~bc1fc8fd8/?referer=https%3A%2F%2Fwww.google.com%2F

Schoonjans, F. (2020). Values of the Chi-squared distribution table. Retrieved July 13, 2020, from https://www.medcalc.org/manual/chi-square-table.php

Solomon, D. J., & Björk Bo-Christer. (2012). A study of open access journals using article processing charges. *Journal of the American Society for Information Science and Technology*, 63(8), 1485–1485. https://doi.org/10.1002/asi.22673.

Sojer, M., & Henkel, J. (2010). Code reuse in open source software development: Quantitative evidence, drivers, and impediments. *Journal of the Association for Information Systems*, *11*(12), 868-901.

Swan, A. (2006). The culture of Open Access: researchers' views and responses. In, Jacobs, Neil (ed.) *Open Access: Key Strategic, Technical and Economic Aspects*. Chandos.

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. (2011) Data Sharing by Scientists: Practices and Perceptions. PLoS ONE,6(6).

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press. *Ch.3*

TU Delft Policy on Open Science (2016), Retrieved from https://www.openaccess.nl/sites/www.openaccess.nl/files/documenten/oa_beleid.pdf

Venkatesh, V., Morris, M., Davis, G.B. and Davis, F.D. (2003). User acceptance of technology: toward a unified view. *MIS Quarterly*, 27(3): 425-478

Vicente-Saez, R., Martinez-Fuentes, C., (2018) Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research,* 88, 428-436. https://doi.org/10.1016/j.jbusres.2017.12.043

Wallis, J., Rolando, E., & Borgman, C. (2013). If we share data, will anyone use them? data sharing and reuse in the long tail of science and technology. *Plos One, 8*(7), 67332. doi:10.1371/journal.pone.0067332

Wang, X., Liu, C., Mao, W., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. Scientometrics *: An International Journal for All Quantitative Aspects of the Science of Science, Communication in Science and Science Policy*, 103(2), 555–564. https://doi.org/10.1007/s11192-015-1547-0

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. MIS Quarterly, 26(2), 13-23.

Willinsky, J. (2005). The unacknowledged convergence of open source, open access, and open science.

Whyte, A., & Pryor, G. (2011). Open Science in Practice: Researcher Perspectives and Participation. *IJDC, 6*, 199-213.

YouHua, C., XueKe, L., & YouFang, C. (2014). Close connections between open science and open-source software. *Environmental Skeptics and Critics*, *3*(3), 47–51. INSERT-MISSING-URL.

Zhu, Y. (2017). Who support open access publishing? gender, discipline, seniority and other factors associated with academics' oa practice. *Scientometrics*, *111*(2), 557–579. https://doi.org/10.1007/s11192-017-2316-z

Zuiderwijk, A., Shinde, R. & Jeng, W. (Accepted), What drives and inhibits researchers to share and use open research data? Proposing the Open Research Data Adoption Model (ORDAM)

Zuiderwijk A, Janssen M, Choenni S, Meijer R. Design principles for improving the process of publishing open data Transforming Government: People, Process and Policy. 2014;8(2):185 - 204.