

Open data sharing in the transportation sector

A quantitative study on the extrinsic factors influencing researchers in the transportation discipline.

Master thesis submitted to Delft University of Technology
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex Systems Engineering and Management**

Faculty of Technology, Policy and Management

by

Emma Charlotte van den Hengel

Student number: 4545834

To be defended in public on March 23th 2023

Graduation committee

Chairperson: Dr. A.M.G. (Anneke) Zuiderwijk- van Eijk, Information and Communication Technology

First Supervisor: Dr. A.M.G. (Anneke) Zuiderwijk- van Eijk, Information and Communication Technology

Second Supervisor: Dr. E.J.E. (Eric) Molin, Transportation and Logistics

Preface

I chose this topic because I followed the Information & Communication track within the Technology, Policy and Management faculty at the TU Delft during my master's studies. However, for my undergraduate studies, at the same faculty, I followed the Transportation & Logistics track. I wanted to combine these two tracks in my master's thesis. With the help of Anneke Zuiderwijk, I succeeded. I would like to thank her for this. For this research, I started in September 2022 and completed my thesis in February 2023.

It has been a long process where I have learned a lot about creating a questionnaire, constructing a thesis, and analyzing so much different data. With the outcome of this study, I hope to contribute to the larger study of openly sharing research data.

For the past 5 months, I have been writing my master's thesis on investigating open research data sharing in the transportation discipline. For this, I have had help from several people. Therefore, I would like to thank the following people.

First of all, I would like to thank my thesis supervisor Dr. A.M.G. Zuiderwijk for her support, feedback and advice. Because of her guidance, I was able to successfully complete my master's thesis.

In addition, my second thesis supervisor provided a very nice supportive role. I could always turn to him for questions about conducting the quantitative research. Therefore, I would also like to thank Dr. E.J.E. Molin very much.

During the testing phase of my questionnaire, I also had help from two researchers at TU Delft's TPM faculty. They gave me fine feedback that gave me better insights into how to put out a questionnaire. Mrs. L. Pilshchikova and Mr. A.E. Abbas thank you for this. I would also like to thank my fellow students who helped me in the process of my master thesis.

In addition, I would like to sincerely thank the respondents of the questionnaire. Without them, I could not have even conducted my research. Thank you again for taking the trouble to fill out my questionnaire.

Finally, I would like to thank my family. They have always motivated and supported me during my master's thesis.

Emma Charlotte van den Hengel

Schiedam, March 23, 2023

Summary

To date, many researchers choose not to share their research data openly. However, the open sharing of research data is becoming increasingly important as more and more data is being generated these days. Sharing research data potentially provides transparency, repetition, and progression (Huettmann, 2015). For several research disciplines, research has already been conducted on researchers' reasons for sharing or not sharing research data and the influencing factors on these researchers. However, within open data research, there is a lack of discipline-specific information on influencing factors for the open sharing of research data. In addition, it is not clear which influencing extrinsic factors are important in the transportation discipline. This causes the development in discipline-specialized incentives and policies to be hindered. It was essential to conduct this study because information was gathered about the influencing factors within this discipline. After gaining this information, it was possible to generate possible developments that might lead to the research data being shared more frequently. Much potential lies within this discipline, as no previous extensive research had been conducted on the extrinsic factors within this transportation discipline. Sharing research data is also very important in this discipline because operational efficiency within this discipline should be as high as possible (Weerakkody et al., 2017).

Therefore, the objective to solve this problem is to identify what the extrinsic factors of researchers in the transportation discipline are and whether these extrinsic factors influence researchers in sharing research data. Previous research has shown that a number of extrinsic factors influence researchers in choosing to share research data openly (Zuiderwijk et al., 2020). A researcher's extrinsic factors focus on the output or impact of an action by the researcher (Beer, 2022). With the knowledge of what extrinsic factors influence academic researchers in the transportation discipline and to what extent, appropriate measures can be devised that will increase the likelihood of openly sharing the research data. The corresponding research question for this study is as follows; *'To what extent do extrinsic factors influence openly sharing research data by academic researchers in the transportation discipline?'*

The methods used to investigate which extrinsic factors influence researchers in the transportation discipline and to what extent are a systematic literature review and a questionnaire. With the narrative literature review, it became clear what extrinsic factors influence researchers with the open sharing of research data. The questionnaire was administered to academic researchers specializing in the transportation discipline to find out which extrinsic factors influence researchers in the transportation discipline and to what extent. 77 respondents completed this questionnaire. After descriptive analysis, 56 respondents remained within the sample for this study. A logistic regression analysis was performed with these 56 respondents and the correlation between the dependent and independent variables was examined.

Through the literature review, this study identified ten extrinsic factors that influence researchers in general when sharing research data openly. The analysis revealed that of these ten extrinsic factors, nine extrinsic factors influence researchers within the transportation discipline. These extrinsic factors are; 'requirements and formal obligations', 'legislation and regulation', 'facilitating conditions', 'social influence and affiliation', 'trust', 'expected performance', 'effort', 'researcher's experience and skills' and 'data characteristics'.

The questionnaire incorporates statements by extrinsic factor. The respondents could indicate to what extent they agreed with these statements. The analysis showed that in the presence of a number of variables, belonging to a specific extrinsic factor, the probability of sharing or not sharing the research data increased or decreased over the past 5 years. The more variables within an extrinsic factor had a relationship with the dependent variable, the more likely it was that the extrinsic factor in question influenced researchers in the transportation discipline to share or not to share research data. It can be concluded that the following extrinsic factors influence researchers in the transportation discipline to share research data openly; 'requirements and formal obligations', 'legislation and regulation', 'facilitating conditions', 'expected performance', 'effort', 'social influence and affiliation', 'researcher's experience and skills' and 'data characteristics'.

The study shows that almost all extrinsic factors that affect researchers in general also affect researchers in the transportation discipline. However, it appears that some extrinsic factors affect the sharing or non-sharing of research data more than other extrinsic factors. An example of an extrinsic factor of greater influence is 'legislation and regulation', which researchers cannot avoid because of the possible consequences of these laws and regulations. The recommendation is to focus primarily on the extrinsic factors that have the most influence on influencing researchers in the transportation discipline and for which measures can be implemented. The extrinsic factors that influence the most are; 'requirements and formal obligation', 'legislation and regulation', 'facilitating conditions', 'researcher's experience and skills', 'expected performance' and 'data characteristics'. Some of the measures that can be taken are; 'researcher's experience and skills' by offering training or courses to enhance experiences. Or, for the extrinsic factor 'facilitating conditions', simplifying the use of data repositories by their developers. These measures can be taken by one of the stakeholders, namely platform developers.

In addition to examining the influencing extrinsic factors, we also examined the barriers to researchers not sharing research data. With this information, it is possible to think of developments that could ensure that more research data is shared. The most common barriers that cause researchers not to share research data correspond to the extrinsic factors mentioned above. For example, the barrier related to the experience of the respondents has a great influence on not sharing the research data. In addition, a strong barrier is that the organization does not want the researcher to share the research data. This barrier again corresponds to the extrinsic factors 'requirements and formal obligations'. Another barrier that strongly affects researchers is that the data contains information that is too sensitive, for which specific laws and regulations have been established. This barrier corresponds to the extrinsic factor 'legislation and regulation'.

The scientific and theoretical contributions are that it contributes to the larger research on open data. This research contributes to this research by investigating extrinsic factors in the transportation discipline. A discipline that has not yet been extensively researched in terms of influencing factors. By conducting this research, the extrinsic factors that influence researchers in the transportation discipline become known and these factors can be included in developments regarding open data within this transportation discipline.

The societal contributions relate to the stakeholders. Each stakeholder has its contributions towards the open data phenomenon. For example, researchers can create more awareness among other researchers to share research data, organizations can promote the sharing of research data, policymakers can possibly take the sharing of research data into account when drawing up laws and regulations, and platform developers can develop the platform in such a way that it attracts more researchers. These contributions try to get researchers to share research data openly more often.

Thus, this research can contribute to the developments that may be taking place within open data research, especially within the transportation discipline. In addition to the recommendations made, it is possible to use this research in making other recommendations that may lead to researchers sharing research data more frequently.

Table of Contents

Preface	2
Summary	3
List of figures	8
List of tables	9
Chapter 1: Introduction	10
1.1 Background information	10
1.2 Problem identification.....	10
1.3 Knowledge gap.....	11
1.4 Research question	11
1.4.1 Sub-questions	11
1.5 Relevance	12
1.5.1 Societal relevance	12
1.5.2 Academic relevance.....	12
1.6 Connection to CoSEM Master program.....	12
1.7 Research approach and methods.....	13
1.8 Structure of the thesis	14
Chapter 2: Theoretical framework	15
2.1 Narrative literature review.....	15
2.1.1 Approach literature review.....	15
2.2 Definitions of key concepts.....	19
2.2.1 Open research data	20
2.2.2 Intrinsic and extrinsic.....	20
2.2.3 Transportation data	21
2.3 Influencing factors on open data sharing	22
2.3.1 The researcher’s background	22
2.3.2 Requirements and formal obligations	23
2.3.3 Personal drivers and intrinsic motivations	23
2.3.4 Facilitating conditions.....	23
2.3.5 Trust.....	24
2.3.6 Expected performance	24
2.3.7 Social influence and affiliation.....	24
2.3.8 Effort.....	25
2.3.9 The researchers’ experience and skills	25
2.3.10 Legislation and regulation	25
2.3.11 Data characteristics	25
2.4 Conclusion systematic literature review.....	26
Chapter 3: Data collection	28

3.1 Conceptual Framework	28
3.2 Questionnaire.....	29
3.3 Questionnaire layout.....	29
3.4 The questionnaire testing process.....	30
3.5 Questionnaire distribution	30
3.6 Questionnaire results	31
3.7 Data analysis techniques.....	31
3.7.1 Principal Components Analysis.....	31
3.7.2 Descriptive Analyses	32
3.7.3 Logistic Regression Analyses.....	32
Chapter 4: Data analysis	34
4.1. PCA for the extrinsic factors.....	34
4.1.1 Requirements and formal obligations	34
4.1.2 Legislation and regulation	35
4.1.3 Facilitating conditions.....	35
4.1.4 Trust.....	36
4.1.5 Expected performance	36
4.1.6 Social influence and affiliation.....	37
4.1.7 Effort.....	38
4.1.8 Researcher’s experience and skills	38
4.1.9 Data characteristics	39
4.2 Logistic Regression Analyses.....	39
4.2.1 Descriptive analyses for the extrinsic factors	39
4.2.2 Logistic Regression Analyses for the extrinsic factors	41
4.2.3 Change within open sharing of research data	42
4.3 Descriptive analyses for the barriers of sharing research data	43
Chapter 5: Discussion and recommendation	45
5.1 Discussion of results.....	45
5.2 Recommendation for the stakeholders.....	47
Chapter 6: Conclusion	49
6.1 Conclusion	49
6.1.1 Conclusion first sub-question	49
6.1.2 Conclusion second sub-question	49
6.1.3 Conclusion third sub-question.....	50
6.1.4 Conclusion research question.....	50
6.2 Scientific and theoretical contributions.....	50
6.3 Societal relevance and contributions	51
6.4 Limitations	52
6.5 Avenues for future research	53

References	54
Appendices	60
<i>Appendix A. Studies about open data sharing per discipline</i>	60
<i>Appendix B. The sources of the extrinsic factors influencing data sharing</i>	62
<i>Appendix C. Narrative Literature Review</i>	66
<i>Appendix D. The complete questionnaire</i>	69
<i>Appendix E. HREC approval</i>	76
<i>Appendix F. Additional results</i>	77
Background information.....	77
Characteristics of the research fields	78
Data sharing characteristics.....	78
Nationality and country of residence of the respondents.....	78
Current positions and the locations of the respondents.....	79
Influencing characteristics per research field on sharing research data	80
Data characteristics	80

List of figures

Figure 1 Phases of the thesis	14
Figure 2 Key words	20
Figure 3 Researchers' drivers of whether to share and use data (Zuiderwijk et al., 2020)	21
Figure 4 Influencing extrinsic factors on data sharing	27
Figure 5 Conceptual framework	28
Figure 6 Conceptual model with existing effects.....	45
Figure 7 HREC approval	76

List of tables

Table 1 Research methods	14
Table 2 Articles literature review	16
Table 3 Aspects regarding the researcher's background.....	23
Table 4 Aspects regarding requirements and formal obligations.....	23
Table 5 Aspects regarding facilitating conditions	24
Table 6 Aspects regarding trust.....	24
Table 7 Aspects regarding expected performance	24
Table 8 Aspects regarding social influence and affiliation.....	25
Table 9 Aspects regarding effort	25
Table 10 Aspects regarding the researchers' experience and skills	25
Table 11 Aspects regarding legislation and regulation.....	25
Table 12 Aspects regarding data characteristics	26
Table 13 PCA for requirements and formal obligations	35
Table 14 PCA for legislation and regulation.....	35
Table 15 PCA for facilitating conditions.....	36
Table 16 PCA for trust.....	36
Table 17 PCA for expected performance.....	37
Table 18 PCA for social influence and affiliation	37
Table 19 PCA for effort	38
Table 20 PCA for researcher's experience and skills.....	38
Table 21 PCA for data characteristics	39
Table 22 Correlation matrix extrinsic factors	40
Table 23 Logistic Regression extrinsic factors	42
Table 24 Number of stimulated respondents who previously did not share	43
Table 25 Descriptive results for the barriers	44
Table 26 Studies about open data sharing per discipline	60
Table 27 The sources of the extrinsic factors influencing data sharing.....	62
Table 28 Narrative Literature Review.....	66
Table 29 Background figures	77
Table 30 The influencing characteristics of the transportation discipline on open data sharing.....	78
Table 31 Data sharing characteristics	78
Table 32 Data sharing characteristics	78
Table 33 Nationalities and countries of residence of the respondents.....	79
Table 34 The academic roles of the respondents per country of living.....	79
Table 35 The influencing characteristics of the research field on open data sharing	80
Table 36 The influencing characteristics of the research field on open data sharing	80
Table 37 The data sizes respondents work with.....	80
Table 38 Distribution of data size relative to data sharing.....	81
Table 39 The data formats respondents work with.....	81
Table 40 Distribution of data format relative to data sharing.....	82
Table 41 The nature of the data respondents work with	83
Table 42 Distribution of data nature relative to data sharing	83
Table 43 Research funded by a third party.....	83
Table 44 Distribution of research funded relative to data sharing.....	84

Chapter 1: Introduction

This introductory chapter first provides background information on the data sharing by researchers. This background information leads to the identification of the problem and the corresponding research question. The social relevance of this research is also explained. As well as the research approach and method. In the last part of this chapter, a structure of the thesis is given.

1.1 Background information

Data has been part of our lives for over 400 years (Campbell, 2015). Data is the principle for science and its further development. However, to develop and improve science, it is important to share the scientific data obtained with others. 'Open data is non-privacy-restricted and non-confidential data which is produced with public money and is made available without any restrictions on its usage or distribution' (Charalabidis et al., 2012, p. 258). Data sharing potentially provides scientific benefits (Zuiderwijk & Spiers, 2019). The open sharing of this data potentially forms multiple perspectives, identifies errors, creates new hypotheses, and uses the data for training purposes (Piwowar & Vision, 2013). Other possible benefits that come from openly sharing research data include reducing fraud and research funding is used efficiently (Piwowar & Vision, 2013). More and more data is being generated and shared (Campbell, 2015). However, there are also reasons why researchers are not sharing their data. This depends on several factors, such as current laws and regulations, researchers are not encouraged to share their data, and there are few good platforms that arrange data thematically so that it is easy for both sharers and users (Campbell, 2015). Some research disciplines are more affected by these factors, preventing data from being shared, than others (Zuiderwijk et al., 2020). A reason for this can be because within disciplines there are different academic cultures, traditions and nuances that lead researchers to share or not share their research data openly. There are research disciplines where data is barely shared with each other, to stay ahead of the competition or ensure security (Kim & Adler, 2015; Yoon & Kim, 2017; Zenk-Möltgen et al., 2018). For example, research disciplines that deal with highly sensitive information, such as health care, wildlife ecology and social science (Zuiderwijk & Spiers, 2019; Zuiderwijk et al., 2020). Research disciplines where it is more common to share research data openly are genetic genealogy, atmospheric science, and oceanography (Zuiderwijk et al., 2020).

For a comprehensive listing of previous studies by discipline, Appendix A can be consulted. This lists all related studies by discipline.

1.2 Problem identification

More and more data is being shared. Yet it seems that many academic researchers decide not to share their research data openly (Zuiderwijk et al. 2020). Several countries are investing in encouraging data sharing that relates to public services, such as transportation, social services, health, education, and crime (Weerakkody et al., 2017). One discipline that has not yet been the subject of much research regarding open data sharing is the transportation discipline. A reason could be that much data regarding this transportation discipline is held by the public or private sector. Until a few years ago, this data was not shared by these agencies and the academic researchers involved who worked with this data (Weerakkody et al., 2017). However, the data is now being shared openly. This makes the data more accessible to everyone, including academic researchers. The academic researchers can also, with the permission of the involved agency, more easily choose to share the research data. It is already known from previous research that there are several extrinsic factors that influence researchers to openly share their research data (Zuiderwijk et al., 2020). The extrinsic factors focus on the output or impact of the researcher sharing the research data (Beer, 2022). The problem involved here is that there is not yet much understanding about the data sharing of researchers in the transportation discipline and whether the extrinsic factors influencing the researchers match those of researchers in other disciplines. **The objective to solve of this study is to identify what the extrinsic factors of researchers in the transportation discipline are and whether these extrinsic factors influence researchers in sharing research data.** Currently, there are limited discipline-specific insights into the extrinsic factors that influence researchers to openly share

research data. This causes the development in discipline-specialized incentives and policies to be hindered. It is important to identify extrinsic factors for the transportation sector so that developments can be developed to allow researchers to share more research data openly. These developments, in turn, encourage that research data is shared openly more often. It is important for researchers within the transportation discipline to openly share research data, as there is much potential within this discipline. This potential lies mainly with this discipline, as there has been no previous extensive research on this discipline. However, it is not yet known what extrinsic factors influence researchers within this discipline and what consequences this influence leads to. It is important to share research data openly as much as possible because it potentially creates transparency, repetition, and progression in the research discipline (Huettmann, 2015). Open sharing of research data creates potentially new information and knowledge within the discipline (Zuiderwijk & Spiers, 2019). Thus, other researchers can move forward with open issues or researchers can jointly find measures to complex problems related to the transportation discipline (Weerakkody et al., 2017).

1.3 Knowledge gap

The literature shows that there has been little research on open data sharing by researchers in the transportation discipline (Mahajan et al. (2021). The qualitative study by Mahajan et al. (2021) shows that data sharing in the academic sectors did increase in recent years. However, this figure is focused on the open data shared by the private and public sectors and not the research data of researchers in the transportation discipline. Mahajan et al. (2021) indicate that previously, private and public organizations often decided not to share transportation data to avoid criticism, ensure user privacy and safeguard corporate interest. Previous research shows that more and more data is being shared, but proportionately much more data is not being shared (Zuiderwijk et al., 2020). Today, more and more data is also being shared within the transportation discipline (Weerakkody et al., 2017). This data is often in the hands of the public or private parties. Researchers often conduct their studies for these parties. Previously, the results of these researchers were only for the parties themselves. However, these parties are sharing their data more often, which means that the researchers are also allowed to share their research data more often. The possibility of researchers sharing research data has, among other things, to do with proprietary rights or agreements made with the relevant parties (Zuiderwijk et al., 2020). These interlocking agreements are a factor that can possibly influence researchers to share research data openly or not. However, there are other extrinsic factors that can influence researchers (Zuiderwijk et al., 2020).

1.4 Research question

Now the question is what specific extrinsic factors and to what extent these factors influence researchers to share their research data. Therefore, the main question for this problem is; *'To what extent do extrinsic factors influence openly sharing research data by academic researchers in the transportation discipline?'*

1.4.1 Sub-questions

The sub-questions will support the research question in finding the conclusion. Each sub-question is further explained in what way that sub-question will be answered. The sub questions are as follows;

- SQ1: *What extrinsic factors influence academic researchers in openly sharing research data?*

This sub-question is answered by conducting a literature review. Multiple platforms are used where this literature search is conducted, such as Google Scholar or Scopus. The method of searching for this literature is conducted using the Narrative Literature Review. The literature is set out in a table so that the search for specific data can be mapped out and followed by the reader. These extrinsic factors are listed and explained so that they can be further worked with during the questionnaire design and data collection.

- SQ2: *What extrinsic factors influence academic researchers in the transportation discipline to share research data openly?*

The second sub-question is answered using a questionnaire distributed to academic researchers and then analyzed through the software IBM SPSS version 29. The presence and frequency of influencing extrinsic factors in the transportation discipline are tested. From this analysis, sub-question two can be answered.

- SQ3: *What are the most common barriers to academic researchers in the transportation discipline to share research data openly?*

The third sub-question is answered by inserting into the questionnaire, questions related to the individual barriers of sharing of open research data. From this data possible measures are derived for example, policy makers to make it easier for researchers to be able to share data or platform developers to make it more convenient for researchers to be able to add data to a platform.

1.5 Relevance

This section describes the societal relevance and the academic relevance. The importance of this study is to clarify what factors influence researchers to share or not share research data. Based on these results, measures can be taken to ensure that research data are shared more often.

1.5.1 Societal relevance

One of the benefits of openly sharing (research) data is that it potentially contributes to transparency within a discipline. This transparency is important for both researchers and citizens. Transparency leads to trust because everyone has insight into the data and can verify for themselves what is in the data (Morey et al., 2015). Open data sharing also helps with analyzing large volumes of data, test newly constructed hypotheses, replicate research, and avoid duplication of studies (Zuiderwijk & Spiers, 2019). These aspects provide new knowledge and information. Previous research has shown that the degree of open data sharing varies by research discipline (Tedesoo et al., 2021; Tenopir et al., 2011; Zuiderwijk & Spiers, 2019). Currently, open research data has already been studied in various disciplines. However, there is still little known about the transportation discipline, while a lot of data is processed in this discipline (Mahajan et al., 2021). Consider GPS systems and traffic flows. When this data becomes more available, it is possible for researchers to analyze these large amounts of data, create new hypotheses from analyzing it and possibly gain more knowledge from it. The knowledge derived from this available data can potentially lead to improved operational efficiency (Weerakkody et al., 2017). In the transportation discipline, operational efficiency can manifest itself in reducing congestion on the road network. Operational efficiency can also relate to public transportation ride times (Weerakkody et al., 2017). This research will be one of the first to focus on the transportation discipline regarding the use of open data sharing. **The objective to solve of this study is to identify what the extrinsic factors of researchers in the transportation discipline are and whether these extrinsic factors influence researchers in sharing research data.** The outcome of this research may influence the future use of open data sharing in the transportation discipline and any additional new information and knowledge that comes from open data sharing.

1.5.2 Academic relevance

The research is also scientifically relevant. Science is based on transparency, repetition, and progression (Huettmann, 2015). Sharing research data openly respects these aspects. The open sharing of research data potentially leads to transparency (Huettmann, 2015). For accurate and certain outcomes, some researchers will conduct studies multiple times. The outcomes are compared, and a conclusion is drawn from them. Sharing the research data potentially allows researchers around the world, without direct contact, to compare when the studies conducted are identical to each other. This contributes to the aspect of repetition. The progression aspect is also affected by openly sharing research data. Open sharing of research data allows other researchers to continue with the research where the previous research ended. It also allows multiple researchers to consider the problem, which can lead to new insights. This contributes to progression in the research field.

1.6 Connection to CoSEM Master program

The problem statement relates to the Complex Systems Engineering and Management Master because it deals with a complex socio-technical environment, namely the open sharing of research data by academic researchers. Within this environment are multiple stakeholders who are affected by the open sharing of research data. These stakeholders of the open data research all have different norms, values, desires, and a

view of this problem. The research on open data sharing focused on institutional components, technical components, and process components. Thus, open data sharing may depend on institutional components that ensure that open data sharing is prevented or hindered. Examples of these institutional components include the laws and regulations governing the privacy of data subjects. Or the agreements established between researchers and third parties. In addition, there are technical and process components associated with open data sharing to ensure that the data gets to the right place. The technical components focus primarily on the aspects that comprise the data repositories on which research data can be shared. The process components focus on the sharing of the research data, as a process. In addition, several methods covered during the CoSEM master study were used within this study. First, a narrative literature review was conducted, to which key concepts were explained. In addition to explaining these key concepts, this literature review was also used to examine existing theories regarding open data research. Based on the literature review, we searched for influencing extrinsic factors on researchers' open sharing of research data. Using these extrinsic factors, an analysis was conducted, testing which factors apply to researchers in the transportation discipline. Next, a quantitative study was conducted using a questionnaire that was administered to the target group, the academic researchers in the transportation discipline. This questionnaire was analyzed using the analysis techniques covered in the master CoSEM, as well as in the bachelor Technology, Policy and Management. These methods are descriptive analyses, principal component analyses and logistic regression analyses.

1.7 Research approach and methods

The research approach chosen is a combination between inductive and deductive quantitative research. A quantitative study is conducted by using statistical and numerical data analysis methods (Taguchi, 2018). The two main methods of analysis that make up a quantitative study are descriptive statistics and inferential statistics. Descriptive statistics provides information about the distribution of data from the questionnaire. The inferential statistics uses techniques that can tell more about the relationships between variables, such as correlations. Both of these methods of analysis are used, however, the inferential statistics adds the depth (Taguchi, 2018). Within this study, there is both inductive and deductive research. The extrinsic factors that influence researchers from certain disciplines in openly sharing research data are already known from previous studies. This gives the research a deductive character. The inductive nature of this research is that the extrinsic factors for the transportation discipline are not yet known and need to be examined (Verhoeven, 2018). Previous research reveals the drivers and inhibitors that influence researchers to openly share research data (Zuiderwijk et al., 2020). Using this literature, additional questions are drawn up for researchers in the transportation discipline. The corresponding research question was discussed earlier in this proposal and reads as follows; *'To what extent do extrinsic factors influence openly sharing research data by academic researchers in the transportation discipline?'* The methods to this research consist of a narrative literature review, a questionnaire, and an evaluation of the questionnaire results. This thesis is therefore divided into different phases of research. The first phase focuses on the research analysis. The research analysis analyses the background information, the problem, the research objectives, the corresponding research question and sub-questions, and the method used in the process. The second phase is the theoretical framework, in which the literature review is conducted. The data collection phase ensures that the questionnaire is created and deployed to collect data. In addition, the forms of analysis used to analyse the data were explained. The fourth phase analysis the data from SPSS, reports the observations and discusses the results. Then the fifth phase answers the research question, and provides a conclusion, discussion, and recommendation. The final stage focuses more on completing the master thesis by preparing a presentation and completing the report.

Figure 1 shows all phases of the study. In addition, the phases in which the sub-questions were addressed are indicated. Naturally, subquestions, along with the research question, are also addressed in the conclusion

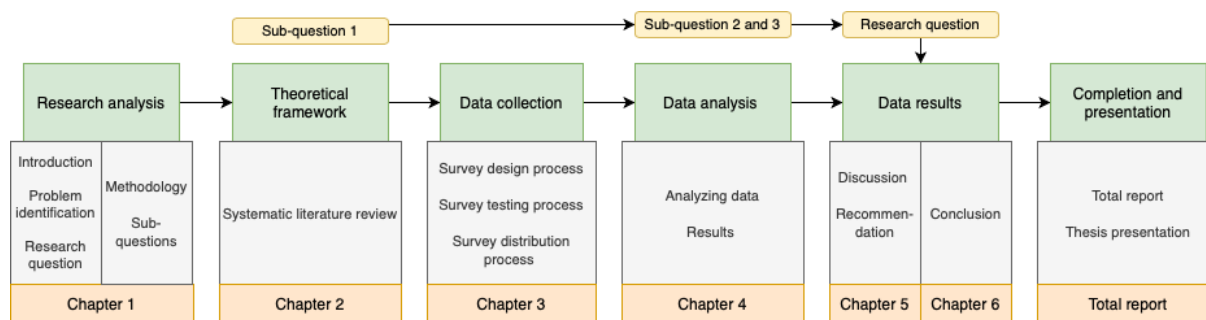


Figure 1 Phases of the thesis

The research methods make clear what data is needed per phase of the research. Based on the research method and the required data it is possible to determine how these data can be provided. The table below shows the research method, the data requirements, and the way to find these data for each sub-question and the corresponding research phase.

Table 1 Research methods

Sub-questions	Research phase	Research methods	Data requirements	Sources/ Tools
What extrinsic factors influence academic researchers in openly sharing research data?	Theoretical framework phase	Narrative Literature Review	Keywords to find literature that helps find answers to the questions.	Literature from Google Scholar and Scopus.
What extrinsic factors influence researchers in the transportation discipline to share research data openly?	Data analysis phase	Questionnaire to gather empirical data	Literature to set up the questionnaire and the data from the questionnaire must be analyzed.	IBM SPSS version 29
What are the most common barriers to academic researchers in the transportation discipline to share research data openly?	Data results phase	Questionnaire to gather empirical data	The analyzed data from the questionnaire and optionally additional literature	Literature from Google Scholar and Scopus and IBM SPSS version 29.

The first sub-question will be answered by a narrative literature review. This literature can be found through a variety of platforms using keywords. The method used for this research is a quantitative study, through a questionnaire. The reason why this method was chosen is that multiple academic researchers in the transportation discipline need to be tested in order to make statements about the factors that influence the researchers in this discipline. The questionnaire is distributed to the researchers and the completed questionnaires are then implemented in the program SPSS that can analyze these data. Based on this analysis, it may be possible to make statements about the influence of this extrinsic factors on the researchers in the transportation discipline. Therefore, the second sub-question will be able to be answered based on the questionnaire results. The third sub-question focuses on the barriers respondents give in the questionnaire to not sharing the survey data. From these results, it is possible to offer advice to policymakers or other stakeholders to increase the likelihood of openly sharing research data.

1.8 Structure of the thesis

The thesis begins with the introduction where the problem is introduced. Then, in Chapter 2, the theoretical framework is elaborated using an extensive and systematic literature review. Chapter 3 explains the approach of this study. Chapter 4 presents the process of setting up the questionnaire. Chapter 5 conducts and describes all analyses. The results of these analyses are presented in tables. The discussion and recommendation are presented in Chapter 6. After which the conclusion follows in Chapter 7. There are also appendixes where additional information can be found.

Chapter 2: Theoretical framework

Several literature reviews exist. One is the narrative literature review. This chapter will cover the narrative literature review. The literature review is an examination of existing information in literature. The literature review summarizes and contextualizes this information. The literature review helps to explain key concepts and the supports answering the first sub-question and the main research question. The main question for this problem is; *'To what extent do extrinsic factors influence openly sharing research data by academic researchers in the transportation discipline?'.* The purpose of this theoretical framework is to answer the first sub-question. By answering this sub-question, it should become clear what extrinsic factors influence researchers' sharing of open research data. The first sub-question is; *'What extrinsic factors influence academic researchers in openly sharing research data?'.*

2.1 Narrative literature review

The narrative review is one of several types of literature reviews. The narrative literature review examines existing information and synthesizes it into a well-flowing and connecting narrative (Green et al., 2006). This literature review describes the findings of the researcher about the already know knowledge or developments of a problem or a situation. The narrative literature review allows findings from multiple articles to be made readable in one study or article. It then further contributes to the conduct of the research. Therefore, it further contributes to the conduct of the research because this literature review supports the research by clustering and summarizing the appropriate information (Green et al., 2006). Therefore, the purpose of a literature review is to evaluate current theories regarding open data sharing. An additional purpose of a literature review is to test current knowledge about open sharing of research data within different disciplines (Baumeister & Leary, 1997). **The objective to solve of this study is to identify what the extrinsic factors of researchers in the transportation discipline are and whether these extrinsic factors influence researchers in sharing research data.** So it is important in the literature review that a distinction has been made between all the influencing factors and the extrinsic factors that influence researchers in choosing to share research data openly.. This literature review first separates the factors. Further in the research, it should be seen whether these extrinsic factors apply specifically to researchers in the transportation discipline.

The method of gathering information in narrative literature review is not systematic, as it may be in other forms of literature review. Nevertheless, there are a couple of aspects within this literature review that are important to write a good narrative literature review (Green et al., 2006). The first aspect is that 'preliminary research' must have been conducted. This gives the researcher information about how many studies have been conducted on the topic in question and in what way. On this information, the researcher can define the problem and the research method within the study can be determined. In addition, this preliminary research ensures that studies are not duplicated. In addition, it is important to use multiple databases from which the literature was drawn (Green et al., 2006). Searching the literature can be done using keywords. These keywords are established based on the problem and topic within the study. To avoid having to study all the literature, it is important to use multiple keywords while searching for literature. These keywords are used as search terms and then the literature is selected according to the occurring keywords and also the title of the literature. The title of the literature also already gives a lot of information about what is being described in the article. Furthermore, it is also important to establish criteria by which the literature is judged to include or exclude articles from the narrative literature review (Green et al., 2006). These criteria can be distinguished into inclusion criteria. These are the criteria that ensure that the literature is used in the literature review. The exclusion criteria eliminate the literature for the literature review.

2.1.1 Approach literature review

By using a narrative literature review, there is no clear system in the literature search for the literature review. However, information was gathered as preliminary research for the problem identification of this study. The preliminary research began with literature from the first supervisor of this study. From this literature, a further search was made using the reference lists within the obtained literature.

Then, after this preliminary research, keywords related to the topic and problem within this research were established. These keywords were found using the literature that was examined in the preliminary research. Keywords were also prepared based on the research questions that were asked. The most common and important terms were written down and used when searching for literature in the Google Scholar and Scopus databases. These keywords can be found in Figure 2 in the next section.

Most of the literature was selected based on the literature obtained from the supervisor and the reference lists within this literature. The other literature was selected based on the occurrence of these keywords in the literature and by the title of the article. It was also chosen not to use too old articles within the literature review. However, no new extensive literature existed for some topics, as literature from years ago could still be adopted. The literature was selected for Dutch and English. Other languages were eliminated.

For this literature review, there are 44 articles used. The table below lists these articles, along with the keywords used to find the article and its publication. Appendix C provides a comprehensive review of the literature used for the literature review along with the study objective, publication and keywords used.

Table 2 Articles literature review

Author	Title and year of publication	Publication	Keywords
Campbell	Access to scientific data in the 21st century: Rationale and illustrative usage rights review. (2015)	Data Science Journal	'What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption'
Janssen, Charalabidis & Zuiderwijk	Benefits, Adoption Barriers and Myths of Open Data and Open Government. (2012)	Information Systems Management	Mail
Kitchin	The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. (2014)	SAGE Publications Ltd.	Open data
Murray-Rust	Open Data in Science. (2008)	Nat Prec	Open data
Piwowar & Vision	Data reuse and the open data citation advantage (2013)	PeerJ	Mail
Zuiderwijk, Shinde & Jeng	What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. (2020)	PLoS ONE	Mail
Weerakkody, Kapoor, Balta, Irani & Dwivedi	Factors influencing user acceptance of public sector big open data. (2017)	Production Planning & Control	Open data sharing
Ryan & Deci	Intrinsic and extrinsic motivations: Classic definitions and new directions. (2000)	Contemporary educational psychology	Intrinsic and extrinsic motivation
Tenkanen, Saarsalmi, Järvi, Salonen & Toivonen	Health research needs more comprehensive accessibility measures: integrating time and transportation modes from open data. (2016)	International journal of health geographics	Transportation domain open data

Mahajan, Kuehnel, Intzevidou, Cantelmo, Moeckel & Antoniou	Data to the people: a review of public and proprietary data for transportation models. (2022)	Transportation reviews	Transportation domain open data government
Zuiderwijk & Spiers	Sharing and re-using open data: A case study of motivations in astrophysics. (2019)	International Journal of Information Management	Mail
Zuiderwijk, Helbig, Gil-Garcia & Janssen	Special Issue on Innovation through Open Data - A Review of the State-of-the-Art and an Emerging Research Agenda: Guest Editors' Introduction. (2014)	Journal of theoretical and applied electronic commerce research	Zuiderwijk open data
Beer	Kwaliteit van combinatiebanen. (2022)	Tijdschrift voor HRM	Extrinsieke en intrinsieke factoren
Baumeister & Leary	Writing Narrative Literature Reviews. (1997)	Review of General Psychology	Narrative literature review
Green, Johnson & Adams	Writing narrative literature reviews for peer-reviewed journals: secrets of the trade.(2006)	Journal of Chiropractic Medicine	Narrative literature review
Roopa & Rani	Questionnaire Designing for a Survey. (2012)	Journal of Indian Orthodontic Society	Questionnaire method
-	Fields of Research (ANZSRC 2020): 3509 Transportation, Logistics and Supply Chains in Datasets - Dimensions. (z.d.)	Dimensions	Research data sizes transportation
-	Publieke verkeers- en transportationdata. (2020).	Talking Traffic.	Transportation data
Schäfer, Pampel, Pfeiffenberger, Dallmeier-Tiessen, Tissari, Darby, Giaretta, Giaretta, Gitmans, Helin, Lambert, Mele, Reilly, Ruiz, Sandberg, Schallier,Schrimpf,Smit, Wilkinson & Wilson	Baseline Report on Drivers and Barriers in Data Sharing. (2011)	Opportunities for Data Exchange (ODE)	Drivers open research data sharing
Corti, Eynden, Bishop & Woollard	Managing and Sharing Research Data: A Guide to Good Practice (Second). (2019)	SAGE Publications Ltd.	Drivers open research data sharing
Feger, Pertiwi & Bonaiuti	Research Data Management Commitment Drivers: An Analysis of Practices, Training, Policies, Infrastructure, and Motivation in Global Agricultural Science. (2022)	Agricultural Research Knowledge	Drivers open research data sharing
European Commission	Transportation themes. Mobility and Transportation. (z.d.)	European Commission	Sub field of transportation

Huettmann	On the Relevance and Moral Impediment of Digital Data Management, Data Sharing, and Public Open Access and Open Source Code in (Tropical) Research: The Rio Convention Revisited Towards Mega Science and Best Professional Research Practices. (2015)	Springer Science+Business Media	Data sharing relevance
Tedersoo, Küngas, Oras, Köster, Eenmaa, Leijen, Pedaste, Raju, Astapova, Lukner, Kogermann & Sepp	Data sharing practices and data availability upon request differ across scientific disciplines. (2021)	Scientific Data	Data sharing disciplines
Tenopir, Allard, Douglass, Aydinoglu, Wu, Read, Manoff & Frame	Data Sharing by Scientists: Practices and Perceptions. (2011)	PLoS ONE	Data sharing disciplines
Morey, Forbath & Scoop	Customer data: Designing for transparency and trust. (2015)	Harvard Business Review	Transparency data
Taguchi	Description and explanation of pragmatic development: Quantitative, qualitative, and mixed methods research. (2018)	System	Quantitative analysis explanation
Zenk-Möltgen, Akdeniz, Katsanidou, Naßhoven & Balaban	Factors influencing the data sharing behavior of researchers in sociology and political science. (2018)	Journal of Documentation	Influence data sharing
Yoon & Kim	Social scientists' data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories. (2017)	Library & Information Science Research	Data sharing disciplines
da Costa & Leite	Factors influencing research data communication on Zika virus: a grounded theory. (2019)	Journal of Documentation	Influencing factors data sharing
Enke,, Thessen, Bach, Bendix, Seeger & Gemeinholzer.	The user's view on biodiversity data sharing— Investigating facts of acceptance and requirements to realize a sustainable use of research data Ecological Informatics. (2012)	Ecological Informatics	Data sharing motivations
Piwowar, Day & Fridsma	Sharing detailed research data is associated with increased citation rate. (2007)	PLoS ONE	Data sharing citation
Ganzevoort, van den Born, Halfman & Turnhout	Sharing biodiversity data: citizen scientists' concerns and motivations. Biodiversity and Conservation. (2017)	Biodiversity and Conservation	Data sharing motivations
Haeusermann, Greshake, Blasimme, Irdam, Richards & Vayena	Open sharing of genomic data: Who does it and why? (2017)	PLoS ONE	Data sharing motivations
Joo, Kim & Kim	An exploratory study of health scientists' data reuse behaviors: Examining attitudinal, social, and resource factors. (2017)	Aslib Journal of Information Management	Data sharing

Kim & Adler	Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. (2015)	International Journal of Information Management	Data sharing motivations
Sá & Grieco	Open data for science, policy, and the public good. (2016)	Review of Policy Research	Open data science
Raffaghelli & Manca	Is there a social life in open data? The case of open data practices in educational technology research. (2019)	Social Media and Open Science	Open data science
Wallis, Rolando & Borgman	If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. (2013)	PLoS ONE	Data sharing infrastructure
Schmidt, Gemeinholzer & Treloar	Open data in global environmental research: The Belmont Forum's open data survey. (2016)	PLoS ONE	Open data science
Zimmerman	Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. (2007)	International Journal on Digital Libraries	Data reuse
Kim & Yoon	Scientists' data reuse behaviors: A multilevel analysis. (2017)	Journal of the Association for Information Science and Technology	Data reuse behavior
Fecher, Friesike & Hebing	What drives academic data sharing? (2015)	PLoS ONE	Data sharing solutions
Ceci	Scientists' attitudes toward data sharing. (1988)	SAGE Publications Ltd.	Data sharing solutions

2.2 Definitions of key concepts

First, some key concepts will be covered. The key concepts are terms that are used frequently throughout the research (Bondi & Scott, 2010). For example, these key concepts are found in the research questions, the problem identification, the questionnaire and in the analysis of the data. The literature review outlined these key concepts so that it is clear what the corresponding definition is for the concept in question throughout the study. The literature found for these key concepts was searched using key words. According to Bondi & Scott (2010), keywords are words that have special status within research. These words play a special role within research. Within this research, these are words that are almost similar to the key concepts. However, these key words were worded a bit more extensively to allow for directional searches of certain literature in the databases. The key concepts for this research are open data, intrinsic and extrinsic, and transportation data.

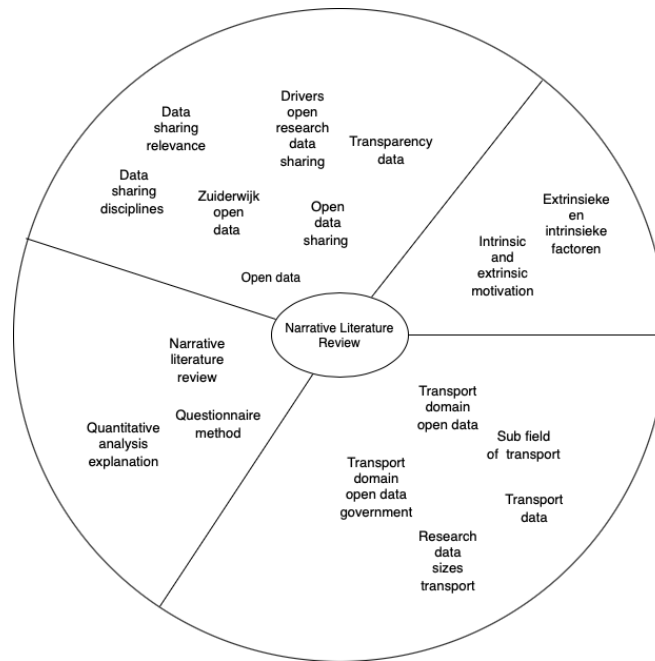


Figure 2 Key words

2.2.1 Open research data

First, what is data? Data comes from the Latin word 'dare' which means 'to give'. However, data is collected through observations, experiments, and other data (Kitchin, 2014). The Latin word for this is 'capere' which should lead to the term 'Capta'. Over time, the term 'Data' has been frequently, wrongly, used instead of 'Capta' and thus it has become the term of elements collected by researchers (Kitchin, 2014). For this study, open data is defined as the following. Open data is data that is available to a specific domain or to anyone. This data can be read, used, and modified without restriction. In addition, this data can be further redistributed with or without restriction (Murray-Rust, 2008). Data exists in many different sectors and branches of society (Kitchin, 2014). This research focuses on data generated by researchers. The scope is therefore the open research data. Open research data is defined as "structured, machine-readable, quantitative data that is published on the Internet and meets all the previously mentioned definitions of data" (Zuiderwijk & Spiers, 2019, p.229). This data should be freely accessible, adaptable, usable, and shareable with other researchers (Zuiderwijk & Spiers, 2019).

2.2.2 Intrinsic and extrinsic

According to the dictionary, the definition of intrinsic is inner, inner value, coming from yourself. The definition of extrinsic is the opposite, coming from outside, outward, not belonging to one's being (Extrinsiek - 4 definities - Encyclo, z.d.). Intrinsic focuses on the content of an aspect. While extrinsic focuses on the output of an aspect (Beer, 2022).

Intrinsic and extrinsic is often linked to motivation of a person to perform an action. According to Ryan & Deci (2000), a person's motivation leads to an action. When a person is energetic and active for a task that person will be motivated. However, when someone is uninspired or has no urge, that person is demotivated. So, there are differences in the amount of motivation, but also in the types of motivation. The different types of motivation depend on the goals and consequences associated with the task at hand. When a task is of greater value to the person, the motivation to bring this task to a desired outcome will also be greater. These two aspects indicate the degree of motivation, and this can differ for everyone (Ryan & Deci, 2000). Intrinsic motivation focuses on an activity that a person does for themselves. You can perform tasks based on fun, interests, or challenges. This form of motivation develops from birth to continually learn and explore new

aspects. This natural development in intrinsic motivation contributes to one's cognitive, social, and physical development, as these developments lead to broadening the knowledge and skills one possesses (Ryan & Deci, 2000). Intrinsic motivation exists within the individual but can also exist between individuals. Extrinsic motivation focuses on tasks performed to achieve a specific outcome. These can be certain outcomes that the individual wants to achieve or meeting the social demands of the environment (Ryan & Deci, 2000).

Extrinsic motivation focuses on the outcome of an activity. Extrinsic motivation can be divided into several categories (Ryan & Deci, 2000). The first category is external regulation. Here, a person is motivated to engage in an activity because of external factors. These can be either positive or negative consequences. This category is in complete contrast to intrinsic motivation. The second category is introjected regulation. This motivation occurs because the person does not want to feel guilt or anxiety and therefore becomes motivated to perform an activity. A person may also become motivated by pursuing a sense of pride. The third category is identification. Here, the person becomes motivated by the higher outcome of the activity and takes the performance of the activity for granted to achieve this higher goal. The last category is integrated regulation, and this category is very similar to intrinsic motivation. However, this is not yet a form of intrinsic motivation because of the instrumental value of the outcome the activity brings and not the pleasure the activity brings. In this category, the activities you perform are similar to your beliefs and needs. (Ryan & Deci, 2000).

During this study, we will primarily focus on the extrinsic factors that influence researchers to share research data. The intrinsic factors will be omitted because these factors are difficult to measure and differ for each individual due to their intrinsic, individual nature. In this study, the extrinsic factors focus more on the output or impact of the researcher sharing the research data (Beer, 2022).

Previous research shows that there are several factors that influence researchers to share or not share research data (Zuiderwijk et al., 2020). This research also shows that researchers become motivated to share their data if it leads to greater recognition and more citations leading back to their work (Zuiderwijk et al., 2020). The figure below shows the most important categories that influence researchers to share or not share research data. On the left side of the figure, the lines represent the influence by the different categories on the researchers' motivation of sharing the research data. On the right side, the lines represent the influence by the categories on using the shared research data. Later in the theoretical framework, a figure will be prepared showing only the extrinsic factors.



Figure 3 Researchers' drivers of whether to share and use data (Zuiderwijk et al., 2020)

2.2.3 Transportation data

With a growing society, there is also an increasing use of all kinds of transportation. Consider motorized and non-motorized vehicles. Because there is so much use of transportation, it is also possible to collect this data (Mahajan et al., 2021). Many different types of data are being generated. Think of traffic flow, GPS systems,

Bluetooth systems, video images and transit data. The increased use of cell phones, affordable sensors, the Internet, and improved communication lines make it possible to generate more data (Mahajan et al., 2021). This data is mainly generated by governments, but also by private organizations. The study by Mahajan et al. (2021) shows that data regarding transportation is little shared. The reason why much data from the transportation discipline is not shared is because of user privacy and competition. Much data generated in the transportation discipline is privacy sensitive (Mahajan et al., 2021). Organizations must comply with the laws and regulations associated with it. When the organization decides to share the data, the degree of privacy is tested so there is no conflict with the laws and regulations (Mahajan et al., 2021). The researchers will regularly conduct the studies for the private or public sector. As such, the researchers will have to deal with the privacy issues that they must adhere to. Examples of sources from which the researchers may obtain their data include traffic signs, matrix signs, parking data, vehicle data, intelligent traffic lights, data for logistics and data from road users (Publieke verkeers- en transportationdata, 2020). Case studies that researchers can focus on are traffic flow, road conditions, speed advisories, parking issues, optimal traffic lights and road details (Publieke verkeers- en transportationdata, 2020).

2.3 Influencing factors on open data sharing

The study by Zuiderwijk et al. (2020) extensively examined other literature on factors influencing researchers when openly sharing research data. These additional sources can be found in Appendix B. literature did not focus on a specific research discipline, but on all kinds of disciplines. The research indicates that these factors influence the drivers or inhibitors of researchers to share their data. Since the number of factors found that influence researchers' sharing of research data is very large, these factors are divided into eleven categories. These categories are:

1. The researcher's background
2. Requirements and formal obligations
3. Personal drivers and intrinsic motivations
4. Facilitating conditions
5. Trust
6. Expected performance
7. Social influence and affiliation
8. Effort
9. The researchers' experience and skills
10. Legislation and regulation
11. Data characteristics

(Zuiderwijk et al., 2020)

Most of the factors influencing researchers to share their research data are in the categories of personal drivers and intrinsic motivations, expected performance and effort to share the research data (Zuiderwijk et al., 2020). However, it is important to consider the large amounts of factors. Not all factors affect all research disciplines. Some factors will not occur at all or to a less extent than other factors (Zuiderwijk et al., 2020).

For this study, it is important to identify extrinsic factors. All intrinsic factors are excluded. For each category, whether these factors have extrinsic or intrinsic influence on the researchers' sharing of the research data will be examined. Thus, the extrinsic factors used in creating the questionnaire are indicated for each category. These factors are shown in the attached tables. As discussed earlier in the theoretical framework, extrinsic factors focus on the outcomes of the action. For example, extrinsic factors may cause research data to be shared because it leads to a desired outcome. However, extrinsic factors can also cause the research data to not be shared precisely because it does not lead to a desired outcome.

2.3.1 The researcher's background

The first category deals with the researcher background (Zuiderwijk et al., 2020). From the research of Zuiderwijk et al. (2020), the factors that fall into this category focus on age of the researcher, seniority in the

research field, gender, and involvement in the research (Zuiderwijk et al., 2020). These are all intrinsic-level factors because they come from within the researcher and do not influence the researcher from the outside. However, there are three factors that do influence the researcher from the outside and these are the culture in which the researcher grew up or is currently located, the country in which the researcher is located, and the research field in which the researcher conducts their research (Zuiderwijk et al., 2020; Schäfer et al., 2011). These are factors that the researcher can influence from the outside in the decision to share or not share the research data.

Table 3 Aspects regarding the researcher's background

Factors used in the questionnaire	Nationality
	Country of residence
	Research field
	Academic role

2.3.2 Requirements and formal obligations

The second category focuses on requirements and formal obligations, as Zuiderwijk et al. (2020) call this category. This is a category of factors that cause researchers to be required to share or not share research data. Consider mandates from parties, such as academic institutions, government agencies or private institutions (Zuiderwijk et al., 2020; Corti et al., 2019). Or funding from one of the institutions (Zuiderwijk et al., 2020; Feger, S. et al., 2022). In addition, compliance with policies and codes of ethics is an important factor (Zuiderwijk et al., 2020; Corti et al., 2019). All these factors are extrinsic in that they externally influence the researcher to share or not share the research data. These guidelines or agreements must be adhered to by the researcher or there may be consequences for the researcher or the research.

Table 4 Aspects regarding requirements and formal obligations

Factors used in the questionnaire	Funding by a third party
	Mandates by a third party
	Compliance by a third party
	Codes of ethics

2.3.3 Personal drivers and intrinsic motivations

The category of personal drivers and intrinsic motivations indicates it (Zuiderwijk et al., 2020). These are all intrinsic factors. These factors all have to do with characteristics that come from within the researcher and are individual to everyone. Examples include feeling responsibility, individual incentives, thoughts toward data sharing, laziness and values and norms that researchers have (Zuiderwijk et al., 2020; Corti et al., 2019).

There are no extrinsic factors in this category. As a result, this category was not included in the questionnaire and there are no questions based on the factors within this category

2.3.4 Facilitating conditions

The facilitating conditions are mainly extrinsic factors that help researchers share research data in an orderly manner. These include technical infrastructure, funding, information systems, data repositories, facilitating platforms, communication, and accessibility (Zuiderwijk et al., 2020; Schäfer et al., 2011; Feger, S. et al., 2022). These factors influence the researcher to share or not share the data.

Table 5 Aspects regarding facilitating conditions

Factors used in the questionnaire	Data repositories / facilitating platforms
	(Technical) infrastructure
	Support by data repositories
	Accessibility

2.3.5 Trust

The category focusing on researcher trust has several different factors (Zuiderwijk et al., 2020). A researcher's trust is based on both intrinsic and extrinsic factors. For example, uncertainties about the accuracy of the research data, misinterpretation, ambiguity, fears of misusing the data and any other aspects that may harm the researcher are intrinsic factors (Zuiderwijk et al., 2020; Corti et al., 2019). Examples of extrinsic factors are transparency, validity, data management, data security and simplifying interpretation of the data (Zuiderwijk et al., 2020).

Table 6 Aspects regarding trust

Factors used in the questionnaire	Transparency
	Data security
	Communication / interpretation
	Validity
	Familiarity

2.3.6 Expected performance

Expected performance is also influenced by several intrinsic and extrinsic factors (Zuiderwijk et al., 2020). Intrinsic factors are the social network of the researcher, manner of citation in the research by the researcher and fears of the researcher arising from publishing the research data (Zuiderwijk et al., 2020). Extrinsic factors are criticism, recognition, reward structures, data visibility, the researcher's profile, collaborations, data accessibility and funding (Zuiderwijk et al., 2020; Schäfer et al., 2011).

Table 7 Aspects regarding expected performance

Factors used in the questionnaire	Data visibility
	Criticism
	Citation / recognition
	Collaboration
	Data accessibility
	Reward system
	Punishment system

2.3.7 Social influence and affiliation

The seventh category is social influence and affiliation (Zuiderwijk et al., 2020). This category focuses primarily on the norms, values and cultures in which the researcher interacts. Most of these factors influence the researcher from both internal and external sources, such as the social pressures the researcher may feel or the social norms and values the researcher must adhere to (Zuiderwijk et al., 2020). The factor of culture can be seen as both intrinsic and extrinsic. However, this factor will have more of the intrinsic nature because the researcher grew up with it. The same goes for the social norms and values that have often been learned since childhood and thus become one's own. Pressure from the outside world, however, is clearly an extrinsic factor.

Table 8 Aspects regarding social influence and affiliation

Factors used in the questionnaire	Pressure from the supervisor
	Pressure from the colleagues
	Pressure from the peers

2.3.8 Effort

Effort has an intrinsic nature. However, it can be influenced by extrinsic factors, causing it to move the researcher to share or not share the research data. Examples of these extrinsic factors are the degree of use of the data, easy use of facilitating platforms, (technical) support, recognition, and time investment (Zuiderwijk et al., 2020; Feger, S. et al., 2022).

Table 9 Aspects regarding effort

Factors used in the questionnaire	Ease of use of facilitating platforms / data repositories
	Time investment
	(Technical) support by a data steward
	Ease of sharing data

2.3.9 The researchers' experience and skills

The researchers' experience and skills category contains mostly intrinsic factors. These factors internally influence the researcher to share or not share the research data. These include expertise, experience in sharing data and knowledge of the researchers (Zuiderwijk et al., 2020; Schäfer et al., 2011). An example of an extrinsic factor is that the researcher received education or training to share the data openly (Zuiderwijk et al., 2020).

Table 10 Aspects regarding the researchers' experience and skills

Factors used in the questionnaire	Experience with data sharing
	Expertise to share data
	Education or training to share data

2.3.10 Legislation and regulation

The category of legislation and regulation is one that consists of extrinsic factors. This category consists of regulations and laws that must be obeyed by researchers. These include laws and regulations that focus on privacy, proper sharing of data, transparency of data, national laws and regulations, rules from possible clients or publishers, contracts, confidentiality, and international agreements (Zuiderwijk et al., 2020; Schäfer et al., 2011; Corti et al., 2019; Feger, S. et al., 2022).

Table 11 Aspects regarding legislation and regulation

Factors used in the questionnaire	Ownership laws and regulations
	Scientific integrity laws and regulations
	Privacy laws and regulations

2.3.11 Data characteristics

The last category of data characteristics, like previous categories, has extrinsic factors that influence researchers to share data. This category is about the characteristics of the data that make researchers share the data or not. The extrinsic factors are the size of the data, the format of the data, the nature of the data, privacy, and data quality (Zuiderwijk et al., 2020; Schäfer et al., 2011).

Table 12 Aspects regarding data characteristics

Factors used in the questionnaire	Data size
	Data format
	Nature of the data
	Data quality

2.4 Conclusion narrative literature review

The first sub-question was answered using a systematic literature review. The first sub-question is as follows; *'What extrinsic factors influence academic researchers in openly sharing research data?'*

In this literature review, a number of keywords were discussed and outlined that add value to the remainder of this research. From this it becomes clear what definition is used throughout the research to define open data. The definition is as follows; open research data is structured, machine-readable, quantitative data that is published on the Internet. Next, the keywords extrinsic and intrinsic are defined. Extrinsic factors focus on tasks performed to achieve a specific outcome. Intrinsic factors come from within and the individual acts with these factors for or from themselves. The keyword transportation explains exactly what is meant by this discipline. Transportation is made possible by the automated and non-automated vehicles on the road. Several examples are given of what researchers can engage in in this discipline. Next, all factors influencing academic researchers in sharing open research data were identified. These factors were divided into extrinsic factors and intrinsic factors. **The objective to solve of this study is to identify what the extrinsic factors of researchers in the transportation discipline are and whether these extrinsic factors influence researchers in sharing research data.** So it is important in the literature review that a distinction has been made between all the influencing factors and the extrinsic factors that influence researchers in choosing to share research data openly.. This literature review first separates the factors. Further in the research, it should be seen whether these extrinsic factors apply specifically to researchers in the transportation discipline. For this study, the intrinsic factors are disregarded, and the focus is on the extrinsic factors that influence academic researchers. The figure below lists the extrinsic factors by category. All of these extrinsic factors were used in the questionnaire. It shows that some categories contain more extrinsic factors than others. One category contains no extrinsic factors and is not reflected in the figure below, namely personal drivers and intrinsic motivations. Only the categories that contain extrinsic factors are considered further in this study. The reason is that extrinsic factors influence the researchers from the outside. These factors will be the same for the researchers. Intrinsic factors differ for each researcher because this has to do with the characteristics of the researcher. However, this introduces too much uncertainty so that unilateral conclusions cannot be drawn from the research.

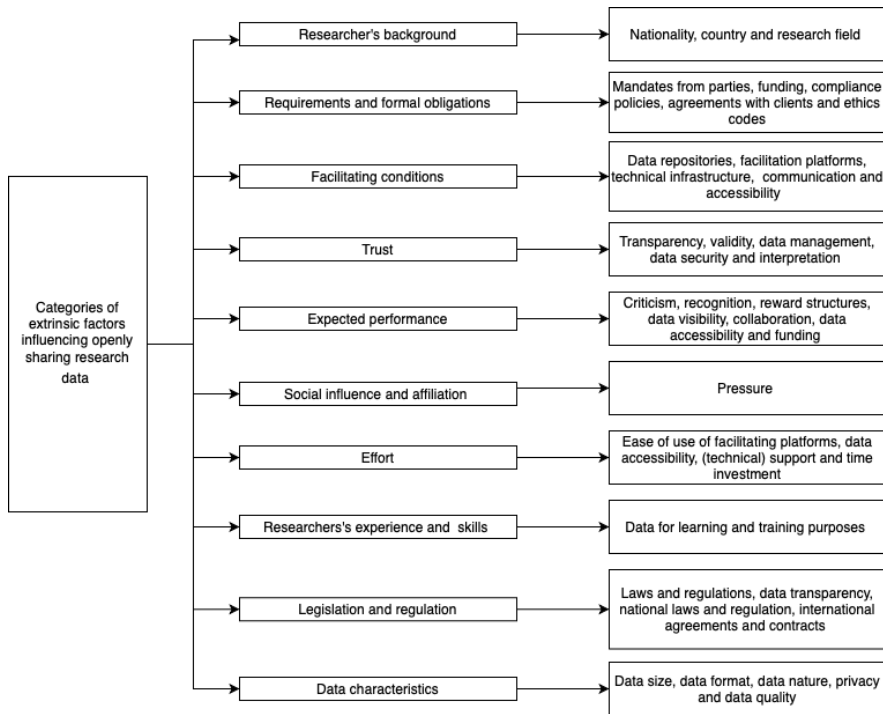


Figure 4 Influencing extrinsic factors on data sharing

In the figure in appendix B, all extrinsic factors are again written out by category along with the source where the factors are explained in more detail.

Chapter 3: Data collection

This chapter explains how the data is collected. The data is collected by creating a questionnaire. This questionnaire is administered to academic researchers focused on the transportation discipline and affiliated with Dutch universities. There are several guidelines to consider when setting up a questionnaire. These guidelines are discussed in section 3.1.

3.1 Conceptual Framework

Figure 5 shows the conceptual framework. This framework shows the relationships between different variables used to answer the sub-questions. The influencing factors are all the factors that generally influence researchers to share or not share research data (Zuiderwijk et al., 2020). In the 'influencing factors' section in figure 5, no distinction has yet been made between intrinsic or extrinsic factors or focused on the transportation research field. Not all influencing factors were analyzed in the analyses. For this study, we are only interested in the extrinsic factors that might influence researchers to openly share research data. Based on the literature review, a distinction was made between intrinsic and extrinsic factors. Thus, it can be noted that the factor 'personal drivers and intrinsic motivations' is not found among the extrinsic factors in the box below it. This is the result of the literature review that was conducted. These extrinsic factors are found in the section 'influencing extrinsic factors' of Figure 5. The relationship between these extrinsic factors toward influencing researchers to openly share research data in the transportation discipline is unknown. The relationships between all extrinsic factors and the dependent variable 'openly sharing research data over the past 5 years' were examined. This relationship was further analyzed in this study.

Next, the possible barriers to sharing research data by researchers in the transportation discipline were also analyzed. The probability of the researchers sharing the research data when certain barriers occur was tested. These barriers can be found on the right section 'Barriers' of Figure 5.

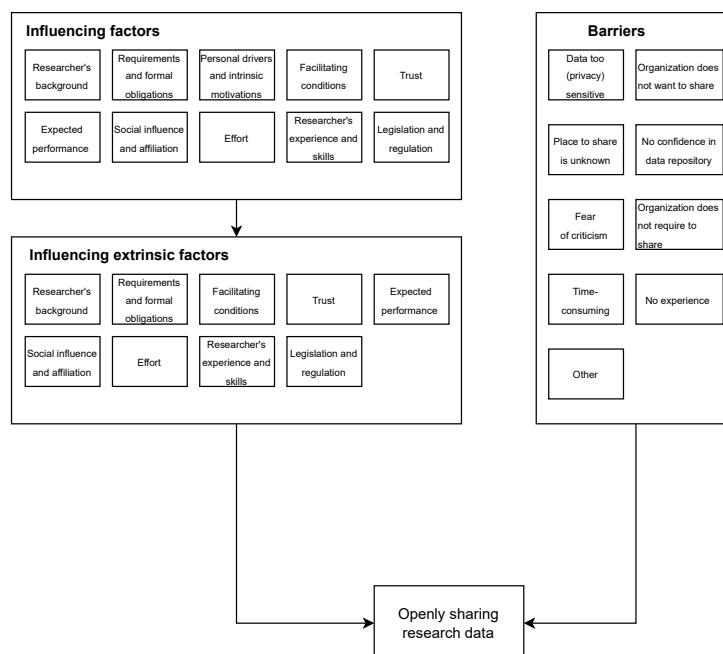


Figure 5 Conceptual framework

3.2 Questionnaire

The data collection methods used in this study is to administer a questionnaire to a group of researchers. To gather good and useful information from the questionnaire, it is important to prepare the questionnaire in the right way. This includes type of questions asked, language used, validity, reliability, and method of distribution (Roopa & Rani, 2012). In the questionnaire related to this study, mainly closed questions are asked, giving academic researchers limited answers to respond to. The answers to the questionnaire questions are based on different scales (Amaresan, 2021). It is important that the questions be simple and clearly worded (Roopa & Rani, 2012). Try to use words that are as neutral as possible regarding their conception (Roopa & Rani, 2012). The reliability is concerned with the results gathered from the questionnaire and whether they are consistent with the research (Roopa & Rani, 2012). Forms of reliability include: test-retest reliability and internal consistency reliability. All these different reliabilities should be checked to test the questionnaire for reliability. The test-retest reliability tests whether the consistency of the answers agree with other answers. Internal consistency reliability tests the reliability of the tool, the questionnaire, itself. Validity deals with the extent to which the questionnaire measures what is necessary to measure for the research (Roopa & Rani, 2012). There are also multiple forms for validity. These forms are; content validity, face validity and criterion validity. Content validity indicates the extent to which the questionnaire reflects all aspects needed for the study. So, from this test it becomes clear whether all important aspects for the research are covered in the tool. Face validity estimates whether the questionnaire is appropriate to administer. This was tested by testing the questionnaire by students and researchers. They went through the questionnaire as a test to see if the questions were interpreted as they should be. Criterion validity indicates the effectiveness of the questionnaire. Do the predicted outcomes match the actual outcomes? There are several methods that can be used to distribute the questionnaire. These include face-to-face interviews, telephone interviews, through the mail or the Internet. For this study, a questionnaire that can be completed via the Internet is used.

These respondents are academic researchers working at a Dutch university focusing on the transportation discipline. The Rathenau institute tracks how many academic staff are employed at Dutch universities. These figures show that about 4560 researchers are employed in the technical sciences sector (*Academische carrière van wetenschappers | Rathenau Instituut, z.d.*). The technical sciences sector can be divided into construction engineering, civil engineering, electrical engineering, industrial design, aerospace engineering, food and agricultural technology, engineering business sciences, transportation, transportation and maritime engineering and mechanical engineering (Meijer, 2018). However, the transportation discipline is a small part in the overarching engineering sciences sector. If it is assumed that all these disciplines are of the same size and the total number of technical researchers can be divided, the result is that each discipline employs about 415 people. However, the disciplines are not of the same size. For example, the transportation discipline will be of a smaller nature than the engineering discipline. However, this does not lead to problems in this study because the population within the transportation discipline is assumed to be larger than it actually is. Thus, a larger sample is also assumed. A larger sample leads to more reliable results. A proper sample requires 200 respondents, considering a 5% margin of error and a 95% confidence level (Steekproefcalculator, 2020). Since there is a time limit, a proper sample will not be achieved. For example, academic researchers take more than a year to conduct their research related to writing a dissertation (Gritter, 2011). The time for this thesis is about one-third of that time and therefore, for plausible results for this master's thesis, we chose to use one-third of the calculated sample. Therefore, the goal is to recruit 60 respondents for this study. However, the result of fewer respondents is that the results are less representative of the entire population.

3.3 Questionnaire layout

The complete outline of the questionnaire can be found in Appendix D. The questionnaire was developed in the Qualtrics software. From Qualtrics the results are exported to the program SPSS where the data is analyzed. The questionnaire begins with an opening statement alerting the respondent to the Informed Consent of the questionnaire. If the respondent agrees, all the respondent has to do is continue with the questions. First, background questions are asked about the respondent that allow the researcher to gain insight into the

respondent. Furthermore, it is important to know if the respondent shares research data and what the characteristics of that data are. The questionnaire is structured so that each extrinsic factor is addressed separately, along with some associated statements.

This structure was chosen to ensure that information about the respondent is known first. By asking for the background information, it became clear which respondents belong to the target group. This structure also ensures that respondents do not have to go through the entire questionnaire before background information is asked, and then the respondent finds out that they do not belong to the target group. This structure also ensures that for respondents who stopped halfway through the questionnaire, it is clear whether or not they belonged to the target group. After this background information, respondents are asked for information regarding the research data with which they are working. This information can all be used for descriptive analysis, to see if the respondents fall within the target group.

Next, respondents are asked to answer several statements divided by extrinsic factor. The statements are divided by an extrinsic factor to give structure and bring clarity to the respondent. They then know what the extrinsic factor focuses on and whether the statements apply to them. It ensures that the statements are placed in the right context by the respondents.

As a final question, the opportunity was given to leave comments or questions. No use was made of this opportunity.

3.4 The questionnaire testing process

The questionnaire was tested by several people to find out if the questions are interpreted correctly, how long the questionnaire takes respondents to complete and if the questionnaire can keep respondents' attention. Four academic researchers tested the questionnaire. Two researchers are supervisors of this study, and the other two researchers were willing to provide feedback regarding the prepared questionnaire. The first supervisor specializes in the field of open data. The second supervisor specializes in the development of advanced stated choice experiments. In addition, the other two researchers specialize in open data and sharing business data. In addition, a student tested this questionnaire on, mainly, the length of the questionnaire. The testers provided feedback to improve the questionnaire.

During a testing appointment, four researchers completed the questionnaire as the target audience should. During this process, feedback was given on any uncertainties or aspects they would address differently. This revealed that some open questions needed to be asked better as closed questions. In addition, it was clearer to adjust the Likert scale to a 5-point scale instead of a 7-point scale. Furthermore, the researchers found the questionnaire to be clear and correctly constructed. The length of the questionnaire is about 15 minutes.

A student in the Engineering and Policy Analysis master's program tested the questionnaire. The student indicated that the questionnaire looked clear. However, it became immediately clear that the student was not the target audience for this questionnaire. Therefore, the student could not comment on the content of the questions in this questionnaire. However, the student came up with details that could be improved, such as font and sentence structure. The only thing the student said is that he found the questionnaire long. Added to that, the student did say that researchers from the discipline probably take more time to complete the questionnaire for a good study and chance for a possible improvement.

3.5 Questionnaire distribution

Distribution of the questionnaire is via e-mail and LinkedIn. Respondents, academic researchers, were approached through the network within the faculty Technology, Policy and Management and the faculty of Civil Engineering and Geosciences at TU Delft. In addition, academic researchers are asked to share the questionnaire with other academic researchers, working at Dutch universities, within the research field. Several researchers shared the questionnaire with their research departments. For example, the questionnaire was

shared by a researcher within the Department of Transportation & Planning at the Civil Engineering faculty. Another researcher at the Civil Engineering faculty directed the researcher of this study to share the questionnaire through the Universities Transportation Study Group mailing list. The questionnaire was sent via email to researchers who signed up to this mailing list.

In addition to mailing the questionnaire to academic researchers, a choice was made to share the questionnaire via LinkedIn. In this post, the supervisors of this master's thesis were named, to include their networks as well. The LinkedIn post was shared on large-scale, leading to a large reach. In addition, some researchers, specialized in transportation discipline, offered to share the questionnaire with their network via LinkedIn. Some academic researchers reposted the researcher's post, others created their own.

The questionnaire was online for 3.5 weeks. The questionnaire was open from Friday, December 2 to Wednesday, December 28. This gives respondents time to complete the questionnaire. In addition, the researcher has time to gather enough respondents.

During these three weeks, respondents were steadily increasing. However, finding academic researchers specializing in the transportation discipline proved difficult. This caused the researcher to take another critical look at the appropriate sample size. The problem lay with the assumption the researcher took regarding the number of academic researchers, with the discipline of transportation, at the faculty of Technology, Policy, and Management. All academic researchers were included in the assumption, while there are also academic researchers in this group who do not specialize in the transportation discipline.

3.6 Questionnaire results

Respondents complete, via the link, the questionnaire in Qualtrics. The researcher has insight into how many respondents completed the questionnaire. After 3.5 weeks on Wednesday, December 28, the researcher used the respondents' answers for analysis. The researcher converted the data to the program IBM SPSS version 29 via Qualtrics. The results analyzed are discussed in the next chapter. During the period the questionnaire was open, 88 respondents responded.

3.7 Data analysis techniques

3.7.1 Principal Components Analysis

During the data analysis phase, the data obtained from the questionnaire is implemented in the program IBM SPSS version 29. In this program, multiple analysis techniques can be applied to obtain the desired results. Normally, factor analysis is applied. However, due to the low number of respondents, there has been chosen to perform principal component analysis. This principal component analysis was chosen because it requires fewer data than the factor analysis. This analysis reduces variables derived from the questionnaire. Reducing the variables from the questionnaire leads to more overview in the analyses. The questionnaire asked about the extrinsic factors that may influence academic researchers regarding the open sharing of research data. These extrinsic factors were elaborated on the basis of various statements with which the respondents could indicate whether they were encouraged with these statements, were neutral in these statements or were discouraged with these statements.

For a better overview, a principal component analysis was performed by extrinsic factor. In this way, when there is an overlap between the statements, the number of statements are aggregated.

First, we looked at how many components remained after performing the principal component analyses (Molin, z.d. -b). These components were determined from the eigenvalues. When the eigenvalues are higher than the value 1, a component is formed. The component with the highest eigenvalue, is the principal component. The remaining variables, or components, with eigenvalues below the value 1 were omitted. Next, it is important to analyze the communality among the variables (Molin, z.d. -b). Communality represents the common variance between the measured variables. In other words, the variance that the variable shares with

the other variables. The rule of thumb for minimum communality is > 0.25 . The component matrix represents the correlation between the component and the variable (Molin, z.d. -b). Here, the higher the values, the more the variable determines the component. All values below < 0.30 are negligible and are not shown in the output. If the communality and values in the component matrix are too low, a choice can be made to exclude the variable. However, it is not necessary to exclude this variable, this can be determined by the researcher. After this analysis, a new factor score was determined by the program SPSS. This factor score can be used again in further analyses. As mentioned earlier in this section, the factor score of the merged component is determined to be used again in further analyses. The factor score is the score of the associated variables for each component. "The factor score is the standardized weighted summation of the standardized values of all variables" (Molin, z.d. -b, p.60).

The basic equation for a factor score is as follows (Molin, z.d. -b).

$$\text{Factorscore} = \text{Zscore}(\sum L_i * Z_i)$$

Where applies;

L_i = Component load of variable i

Z_i = Standardized score of variable i

The load can be obtained from the tables given in the next section. This load can be entered into the equation and along with the appropriate standardizes score, the appropriate factor score for the component in question can be calculated.

3.7.2 Descriptive Analyses

Before logistic regression was applied, a correlation matrix was created. This matrix provides initial insights into the correlations between the independent variables and the dependent variable. For the correlation matrix, there is perfect correlation at the value '1' and no correlation at the value '0'. The higher the value, the more correlation exists between the variables. This correlation matrix is part of a descriptive analysis.

For the third sub-question, only a descriptive analysis was conducted because performing a logistic regression analysis was not possible. This is because only the respondents who have not shared their research data in the past 5 years were asked what their barriers are to not sharing the research data. Since there is a single group of respondents, with the same answer to the dependent variable, it is not possible to perform the regression analysis. For this descriptive analysis, which concerns the barriers leading to not sharing the research data, several dummy variables were also set up. First, the dummy variable 'NotSharing' was set up. This variable indicates with value '1' if the respondents have not shared the research data in the past 5 years. The value '0' indicates if the respondents did share the survey data in the past 5 years. Thus, for this descriptive analysis, the dummy variable is different, so it is easier to interpret during analyzing the positive and negative influences. Next, a dummy variable was created from all barrier possibilities, with the value '1' when the respondents perceived this barrier as a barrier to openly sharing the research data. And the value '0' indicates the remaining respondents, the respondents who do not experience this possible barrier as a barrier but still have not shared research data in the past 5 years, and the respondents who have occasionally shared the research data openly in the past 5 years.

3.7.3 Logistic Regression Analyses

A logistic regression analysis was performed in this section. Logistic regression was used to examine the probability that the dependent variable is determined by the independent variables (Molin, z.d. -a). To perform logistic regression, the dependent variable is a binomial variable and thus can only take one of two values. The dependent variable in this study is 'sharing research data over the past 5 years'. The categories for this are 'yes' and 'no'. This yields a binomial logistic regression analysis. The dependent variable 'SharingData' is a dummy variable where the value '1' is associated with the answer 'yes' and the value '0' is associated with the answer

'no'. The independent variables are those created by performing the principal component analysis. The independent variables can be either categorical or continuous in the logistic regression.

The model fit was determined by first considering the McFadden R Square Test. The McFadden R square test was used to see how the model predicts with the independent variables added. The closer this value is to 1, the better the model predicts. The equation for the McFadden R Square test is;

$$McFadden R^2 = \frac{(-2LL_0) - (-2LL_M)}{-2LL_0}$$

Where applies;

LL₀ = LogLikelihood of the null model

LL_M = LogLikelihood of the fitted model

To determine the relationship between the variables, a number of statistics were looked at. The significance of the Wald statistic tests whether the independent variables affect the dependent variable, sharing or not sharing research data over the past 5 years. However, Null Hypothesis Significance Testing theory presents some problems, mainly when the sample size is not large (McShane et al., 2019). For example, it appears that the null hypothesis is often wrongly assumed based on the p-value coming from the tests. When the null hypothesis is assumed, it is stated that there are no effects between the two tested aspects. However, effects are often small and varying. The small and varying variables are not considered when adopting or rejecting the null hypothesis. According to McShane et al. (2019), it is recommended to interpret the p-value as a continuous variable. The lower the p-value the stronger the evidence that can be drawn from the test (McShane et al., 2019). After looking at the significance within the variables, using a continuous p-value, we looked at exactly what effect the independent variable has on the dependent variable. The effect of the independent variable or the dependent variable is expressed as the coefficient B in the results section. A positive coefficient B shows an increase, while a negative coefficient B shows a decrease within the model. In other words, this coefficient indicates the effect of the independent variable on the dependent variable. This effect can be positive or negative. With this coefficient, the equation of the Logistic Regression Model can be formulated. This model represents the chance of the dependent variable with the independent variables as predictors (Sperandei, 2014). "There is a difference between odds and probabilities. First, probabilities are the ratios between the desired outcomes and the total number of outcomes specified for the independent variables. Whereas odds are the ratios between the probabilities of a desired outcome and the probabilities of the opposite outcome specified for the independent variables. In addition, probabilities are given in values between 0 and 1 and odds from the value of 0" (Sperandei, 2014, p.15). The odds ratio represents the increase or decrease in the odds on the dependent variable by a given factor. The odds ratio is the exponent of the value B, or e^b.

The basic equation for the Logistic Regression Model is as follows (Molin, z.d. -a).

$$\ln\left(\frac{P}{1-P}\right) = a + bX$$

Where applies;

a = Regression intercept (constant) of the dependent variable. In other words, the logit when the value of the predictor is zero.

b = For each unit increase in the predictor, the logit increases by the value b.

For establishing an equation of the logistic regression model, coefficient a is the value between the crossing of column B and the row Constant in the tables given in the results section. The coefficient b can be found back in column B crossed with the relevant independent variable.

Chapter 4: Data analysis

This chapter analyzed the results resulting from the questionnaire using the program SPSS. With these results, it is important to find answers to the second and the third sub-question posed for this study. This second sub-question is as follows; *'What extrinsic factors influence researchers in the transportation discipline to share research data openly?'*. The third sub-question is as follows; *'What are the most common barriers to academic researchers in the transportation discipline to share research data openly?'*

As discussed earlier, 88 respondents completed the questionnaire. Looking at the dataset in more detail, it became clear that there are several empty responses in the dataset. These have been removed from the dataset as this does not contribute to the study. That leaves 80 respondents who at least answered the first question. However, there are also respondents who stopped after several questions. A distinction was made between respondents who completed only the first two questions and respondents who completed more questions. Respondents who only completed the first two questions were not included in the study. The remaining respondents were included because this data may be of interest. Of the 80 respondents, 77 remained. Next, descriptive analysis was used to find out which respondents fell into the target group. Further analyses are conducted with these respondents. This took into account the functions of the respondents. The respondents should see academic researchers in the transportation discipline. All other respondents were removed from the dataset. For more detailed explanations on the distribution of respondents, please refer to Appendix E.

4.1. PCA for the extrinsic factors

4.1.1 Requirements and formal obligations

The table below shows that the four statements can be combined into two components. All the communalities have a value above 0.25. In addition, all values in the components matrix are higher than the value 0.3. So the principal component analysis is performed for this extrinsic factor. There are two components with eigenvalues above the value 1. The first component explains 47.6% of the variance. While the second component explains 26.0% of the variance. The remaining 26.3% is lost.

For easier interpretation, a simple structure was sought within the component matrix. This was achieved as best as possible by orthogonal rotating the matrix. In the table below, a simple structure is presented. However, it is not the perfect simple structure because a double load occurs with one of the variables. However, this table allows for better interpretation of the values. For example, component 1 is labeled requirements by third parties. This component consists of the statements that focus on the influence by third parties on researchers when openly sharing research data. The statement 'funding of my research by funding agencies' determines the label with the largest load. And component 2 is labeled 'requirements by research field'. The second component focuses on the requirements or aspects that apply within the research field. Often journals and codes of ethics are discipline specific, so the second component indicates the extent to which these aspects affect the researcher when openly sharing research data. For this component is the statement 'the encouragement by journals to openly share my research data' the statement with the largest load.

Table 13 PCA for requirements and formal obligations

Requirements and formal obligations	Rotated Component Matrix	
	Component 1 'Requirements by third parties'	Component 2 'Requirements by research field'
Funding of my research by funding agencies	0.870	
Mandates and compliances policies by third parties I collaborate with	0.789	
The encouragement by journals to openly share my research data		0.914
Clear codes of ethics within my research field	0.379	0.748

4.1.2 Legislation and regulation

All the communalities are higher than 0.25 and the correlations from the component matrix are also all higher than 0.3. This ensures that the principal component analysis is performed. There is one component with the eigenvalue higher than the value 1. So, there is 1 component from this principal component analysis. This component explains 64.9% of the variance.

Due to the fact that only 1 component remains after these analyses, the component matrix cannot be rotated, nor does it provide added value. The label this associated with this component is 'laws and regulations'. This component focuses on the laws and regulations that apply to the researcher in the transportation discipline regarding the open sharing of research data. Ownership laws and regulation largely determine the label of this component.

Table 14 PCA for legislation and regulation

Legislation and regulation	Component Matrix
	Component 1 'Laws and regulations'
Ownership laws and regulations	0.894
Scientific integrity laws and regulations	0.739
Privacy laws and regulations	0.809
The requirements to create a Research Data Management Plan	0.772

4.1.3 Facilitating conditions

All communalities do exceed the value 0.25 and all values in the component matrix also exceed 0.30. This ensures that the principal component analysis is performed. The analysis shows that 1 component has an eigenvalue higher than the value 1. This component explains 80.1% of the variances.

The same is true, as in the analysis of the previous extrinsic factor, that no rotating component matrix can be formed by having a single component. All statements have to do with using the data repository. This causes the label of this component to be as follows 'facilitating conditions within data repositories'. All statements have high charges that strongly define the label. However, the load of the statement 'support provided by research data repositories' is the highest.

Table 15 PCA for facilitating conditions

Facilitating conditions	Component Matrix
	Component 1 'Facilitating conditions within data repositories'
Accessible research data repositories	0.819
A supportive (technical) infrastructure or data repositories	0.900
Support provided by research data repositories	0.960

4.1.4 Trust

All communalities are higher than the value of 0.25. In addition, to loads of components are also higher than 0.3. The analysis reveals that there is one component with an eigenvalue higher than the value of 1. This component explains 64.8% of the variances.

As in the previous two extrinsic factors, there is one component and this ensures that no orthogonal rotating component matrix can be formed. This component matrix already provides clarity and easy interpretation. The label for this component is 'trust factors within data repositories'. This component contains statements that contribute to the researcher's confidence in the data repository to possibly share research data there. In this single-component extrinsic factor, the statement 'a data repository using data standards that are accepted in my field' has the largest load. Therefore, this statement contributes the most to the label.

Table 16 PCA for trust

Trust	Component Matrix
	Component 1 'Trust factors within data repositories'
A data repository that provides information about various quality indicators of the data	0.733
A data repository that provides sufficient metadata to allow the interpretation of the data	0.847
A data repository that allows data providers and data users to interact with each other	0.782
A data repository that guarantees the storage and availability of the data for at least the next 10 years	0.843
A data repository provided by one or more organizations I know	0.770
A data repository using data standards that are accepted in my field	0.848

4.1.5 Expected performance

Also for this extrinsic factor, all communalities are higher than the value 0.25. In addition, all component charges are also higher than the value 0.3. This leads to the principal component analysis being performed. The principal component analysis shows that two components can be formed. The first component explains the variances for 51.1% and the second component for 21.4%.

To easily interpret the components and their associated loads, it is important to represent the component matrix as a simple structure. By rotating the matrix orthogonally, this simple structure is obtained. However, the perfect simple structure is not represented because for some variables a double load is known. However, this simple structure provides a better interpretation than the normal component matrix. The label for the first component reads as follows; connection with others. This component consists of statements that focus on

some form of relationship with others, such as collaborations, citations, usability, and visibility. The second component is labeled 'consequences by sharing'. This component consists of two statements that have different consequences. One positive, the other negative. These statements would be expected to correlate oppositely with each other because the consequences associated with them are opposite. One reason for this positive correlation could be that respondents do not get stimulated by any kind of consequence after openly sharing research data and thus these statements have the same effect of stimulating respondents. In the first component, the statement 'the possibility that my research would be cited more often' to the strongest extent determines the label. While for the second component, this is the statement 'a punishment I could receive for not openly sharing my research data'.

Table 17 PCA for expected performance

Expected performance	Rotated Component Matrix	
	Component 1 'Connection with others'	Component 2 'Consequences by sharing'
The possibility that my research will more visible to a wider audience	0.865	
The possibility that my research would be criticized	0.485	0.488
The possibility that my research would be cited more often	0.890	
The data is potentially useful for other researchers	0.876	
Collaboration with other researchers	0.878	
A reward I could receive for openly sharing research data		0.765
A punishment I could receive for not openly sharing my research data		0.871

4.1.6 Social influence and affiliation

The requirements of the communalities and charges within the component matrix have been met. Principal component analyses have been performed. The analysis shows that there is a component with the eigenvalue above the value one. This component explains 78.4% of the variances.

Due to the fact that there is a component, orthogonal rotation is not possible. The component matrix has the correct structure for proper interpretation. The label associated with the component is 'Social norm'. All statements within this component focus on getting advice, although from different people. The statement that to the strongest extent determines the label is 'if my colleagues would tell me I should openly share my research data'.

Table 18 PCA for social influence and affiliation

Social influence and affiliation	Component Matrix
	Component 1 'Social norm'
If my supervisor would tell me I should openly share my research data	0.836
If my colleagues would tell me I should openly share my research data	0.926
If peers from my field would tell me I should openly share my research data	0.893

4.1.7 Effort

For all communalities, all values are greater than the value 0.25. For the loads in the component matrix, all values are higher than the value 0.3. The principal component analyses show that two components have eigenvalues higher than the value one. The first component explains 51.6% of the variances and the second component explains 29.1%.

For analyses with two components or more with the eigenvalue above the value one, where the component matrix is difficult to interpret, an orthogonal rotation was performed. A simple structure was achieved with this rotation. In fact, for this analysis, a perfect simple structure was achieved because no double loads exist in the rotating component matrix. The label of the first component is 'effort' and the label of the second component is 'use and support of data repositories'. The first component is concerned with implementing the open sharing of research data on data repositories. The second component focuses on the use and possible support of the data repository. The statement 'the considerable effort it costs to openly share research data' has the highest load in determining the label. The statement in the second component 'easy to use repositories for openly sharing research data' affects the label the most within the component.

Table 19 PCA for effort

Effort	Rotated Component Matrix	
	Component 1 'Effort'	Component 2 'Use and support of data repositories'
Easy to use repositories for openly sharing research data		0.838
Considerable time that needs to be invested to openly share research data	0.940	
The considerable effort it costs to openly share research data	0.972	
A data steward whom I can ask questions about openly sharing research data		0.784

4.1.8 Researcher's experience and skills

All conditions are met to perform principal component analysis. There is one component with the eigenvalue higher than the value one. This variable explains 80.8% of the variances within the component.

The table below shows the component matrix. Again, rotating the model is not necessary and cannot be performed. The label associated with this component is 'Experience and skills'. All statements attached in this single component have to do with the respondent's current experience or skills and how these experience or skills encourage the respondent to openly share the research data. The statement that determines this label as strongest is 'me having the skills to openly share research data'.

Table 20 PCA for researcher's experience and skills

Researcher's experience and skills	Component Matrix
	Component 1 'Experience and skills'
Experience with openly sharing research data	0.907
Me having the skills to openly share research data	0.939
Me having received training or education on how to openly share research data	0.849

4.1.9 Data characteristics

Both the communalities and the load within the component matrix have values that are all higher than the minimum values. The principal component analysis shows that there are two components with eigenvalues higher than the value one. This means that there are two components that, distributed, contain all statements. The first component explains 43.4% of the variances. The second component explains for 18.4%.

For the analysis of these two components, the component matrix was rotated orthogonally. This rotation allows for easier interpretation. The label for the first component is 'data characteristics' and for the second component this label is 'low quality data'. The first component focuses on various aspects of data characteristics, such as data size, data format, data nature, and high data quality. The other component focuses only on low data quality. It can also be seen that the statement regarding high quality and the statement regarding low quality are oppositely correlated for the second component. This was expected because there is an opposite effect between these two statements. Thus, it was expected that data that the researcher already knows is of low quality is less likely to be shared than when it is of high quality. This is a possible reason for the statement regarding low data quality being included in a separate component from all other statements within the data characteristics. These other statements do not directly give reason to expect the researcher to be less likely to share the research data based on the statement in question. The statement within the first component that contributes most strongly to the component's label is 'my research data being quantitative'. For the second component, this is the statement 'my research being low quality'.

Table 21 PCA for data characteristics

Data characteristics	Rotated Component Matrix	
	Component 1 'Data characteristics'	Component 2 'Low quality data'
My research data being a large data size	0.677	
My research data being a small data size	0.778	
My research data being stored in a commonly used format in my field	0.580	
My research data being qualitative	0.758	
My research data being quantitative	0.794	
My research data being high quality	0.627	-0.593
My research data being low quality		0.935

4.2 Logistic Regression Analyses

As discussed earlier, a logistic regression analysis consists of a dependent categorical variable and multiple independent variables. The independent variables used were developed in principal component analysis. These components are the new variables. First, the effects of the independent variables on the dependent variable were examined, along with the intercorrelation of the independent variables to each other. Next, logistic regression was performed. From this analysis, the mutual effects of all the independent variables can become clear with respect to the dependent variable. In addition to the analysis of the effect of the independent variable, which relate to extrinsic factors, on the dependent variable, an analysis was also made of what most common factors prevent researchers from sharing the research data. Also called the barriers of openly sharing research data in this study

4.2.1 Descriptive analyses for the extrinsic factors

To get a first idea of the interrelationships between the independent variables and the dependent variable 'Sharing data', a correlation matrix was created. Due to the dichotomous nature of the dependent variable, this

is possible. The table below shows the correlations between all variables. The first column shows the correlations with all independent variables. Three independent variables are negatively related to the dependent variable. These are the variables; 'requirements by research field', 'consequences by sharing' and 'low quality data'. This means that due to the presence of the relevant independent variable, fewer researchers shared the survey data over the past 5 years. The correlation matrix shows that the independent variable 'requirements by third parties' has the strongest correlation with the dependent variable. Looking at the correlation table, the variable 'requirements by research field' and 'trust factors within data repositories' have minimal correlation with the dependent variable.

Table 22 Correlation matrix extrinsic factors

Correlations														
	Sharing data	Requirements by third parties	Requirements by research field	Laws and regulations	Facilitating conditions within data repositories	Trust factors within data repositories	Connection with others	Consequences by sharing	Social norm	Effort	Use and support of data repositories	Experience and skills	Data characteristics'	Low quality data
Sharing data	1.000	0.321	-0.004	0.133	0.267	0.000	0.034	-0.175	0.041	0.072	0.095	0.146	0.035	-0.267
Requirements by third parties	0.321	1.000	0.000	0.465	0.069	0.094	-0.221	0.027	-0.207	0.267	0.009	-0.025	-0.105	-0.042
Requirements by research field	-0.004	0.000	1.000	0.294	0.415	0.582	0.358	0.215	0.461	0.111	0.299	0.194	0.418	-0.242
Laws and regulations	0.133	0.465	0.294	1.000	0.374	0.387	0.095	0.336	-0.081	0.288	0.018	0.393	0.287	-0.151
Facilitating conditions within data repositories	0.267	0.069	0.415	0.374	1.000	0.464	0.259	-0.006	0.220	0.071	0.321	0.180	0.223	-0.241
Trust factors within data repositories	0.000	0.094	0.582	0.387	0.464	1.000	0.380	0.308	0.364	0.172	0.389	0.294	0.554	-0.100
Connection with others	0.034	-0.221	0.358	0.095	0.259	0.380	1.000	0.000	0.477	0.119	0.604	0.613	0.524	-0.204
Consequences by sharing	-0.175	0.027	0.215	0.336	-0.006	0.308	0.000	1.000	0.060	0.142	0.160	0.198	0.360	0.229
Social norm	0.041	-0.207	0.461	-0.081	0.220	0.364	0.477	0.060	1.000	-0.176	0.479	0.202	0.464	0.008
Effort	0.072	0.267	0.111	0.288	0.071	0.172	0.119	0.142	-0.176	1.000	0.000	0.197	0.226	0.195
Use and support of data repositories	0.095	0.009	0.299	0.018	0.321	0.389	0.604	0.160	0.479	0.000	1.000	0.495	0.482	-0.195
Experience and skills	0.146	-0.025	0.194	0.393	0.180	0.294	0.613	0.198	0.202	0.197	0.495	1.000	0.318	-0.180
Data characteristics'	0.035	-0.105	0.418	0.287	0.223	0.554	0.524	0.360	0.464	0.226	0.482	0.318	1.000	0.000
Low quality data	-0.267	-0.042	-0.242	-0.151	-0.241	-0.100	-0.204	0.229	0.008	0.195	-0.195	-0.180	0.000	1.000

4.2.2 Logistic Regression Analyses for the extrinsic factors

This logistic regression provides the results that will answer the second sub-question. Logistic regression was used to analyze how the dependent variables were related to the predictors or independent variables (Molin, z.d. -a). Table 22 shows the probabilities of research data sharing by researchers over the past 5 years according to several influencing independent variables. The model represents the influence on the dependent variable by all independent variables.

The way of interpreting these values is explained using the following values. The results show that the variable 'requirements by third parties' has an B coefficient of 1.657. This means for a unit higher score on the independent variable, or 1 standard deviation, it gives a positive effect on whether research data were shared over the past 5 years. In addition, the odds ratio with a value of 5.243 indicates that the odds for choosing to share research data openly over the past 5 years have increased by a factor of 5.243. In other words, the odds of respondents openly sharing research data increased when respondents were influenced by the presence of the independent variable 'requirements by third parties'. For the variable 'laws and regulations', there is a B coefficient of -0.985. This means for a unit higher score on the independent variable, it gives a negative effect on sharing research data over the past 5 years. The odds ratio with a value of 0.373 indicates that the odds for choosing to openly share research data has decreased by a factor of 0.373 over the past 5 years. For the presence of this independent

All other variables can also be interpreted this way. The variables with a negative B coefficient and thus describing a negative effect on open sharing of research data over the past 5 years are 'laws and regulations', 'connection with others', 'consequences by sharing', 'use and support of data sharing repositories' and 'low quality data'. The corresponding odds ratio indicate that the odds for choosing to share research data openly decreases by the corresponding factor. For a negative effect, the closer the value is to the value 0, the stronger the negative effect is on openly sharing the research data. So here there is the possibility that the respondent does not share the research data openly. The strongest negative effects are found with the variables 'laws and regulations', 'connection with others', 'use and support of data repositories' and 'low quality data'. These factors can be found in Table 22.

For the two variables where it is found that they do not correlate with the dependent variables, it also appears that there is a high p-value when running a logistic regression. These variables are 'requirements by research field' and 'trust factors within data repositories'. Thus, for this study, it can be assumed that when respondents are related to these variables, there is no influence on the open sharing of the research data.

The remaining variables have a positive B coefficient, leading to a positive effect on openly sharing research data over the past 5 years. The corresponding odds ratios indicate the odds of choosing to share research data openly increase by the corresponding factor. The higher the odds ratio is, the greater the odds are on sharing the research data openly. Thus, the independent variables 'requirements by third parties', 'facilitating conditions with data repositories', 'experience and skills' and 'data characteristics' have a strong positive effect on the open sharing of research data.

The logistic regression fulfilled a number of expectations for the model. For example, this is true for the variables 'Requirements by third parties', 'Laws and regulations', 'Facilitating conditions within data repositories', 'Effort', 'Experience and skills' and 'Consequences by sharing'. For the variable 'Requirements by third parties', the positive effect is a logical one because these third parties often indicate to the researchers that they want the research data to be shared. It is part of the mutual agreement made. The variable 'Laws and regulations' has a negative effect on the sharing of research data over the past 5 years. This effect is also expected because researchers are less likely to share their research data when there are many laws and regulations that make it difficult to share the research data. Also, the variables related to the ease of sharing

the research data, namely 'Facilitating conditions within data repositories', 'Effort' and 'Experience and skills', all have a positive effect on the dependent variable. This is a logical effect, because if it is made easier for the researchers through, for example, a good data repository or experience in sharing data, that they are also more likely to share the research data. As the last variable 'Consequences by sharing' this outcome is also as expected. When sharing research data carries consequences, especially negative ones, researchers are less likely to share research data. This was also shown in the results.

Table 23 Logistic Regression extrinsic factors

Model			
Predictors	B coefficient	Odds Ratio (Exp(B))	P-value
Intercept (Constant)	-0.307	0.735	0.443
Requirements by third parties	1.657	5.243	0.006
Requirements by research field	-0.349	0.705	0.558
Laws and regulations	-0.985	0.373	0.117
Facilitating conditions within data repositories	2.066	7.893	0.012
Trust factors within data repositories	0.290	1.337	0.625
Connection with others	-2.762	0.063	0.032
Consequences by sharing	-0.485	0.616	0.449
Social norm	0.726	2.066	0.271
Effort	0.522	1.686	0.375
Use and support of data repositories	-0.916	0.400	0.231
Experience and skills	1.620	5.055	0.026
Data characteristics	1.204	3.335	0.122
Low quality data	-1.297	0.273	0.053
McFadden R Square	0.396		

4.2.3 Change within open sharing of research data

In addition, a descriptive analysis was added to provide insight into what independent variable encouraged respondents who previously did not openly share their research data to possibly share research data in the future. Analyzed was how many respondents who have not shared research data in the past 5 years are encouraged to share research data by the given statements. These statements were converted to these existing independent variables in the Principal Components Analysis. This is an assumption made by the researcher of this study and does not mean that fully matches the actual expectation. Thus, the assumption that the researcher makes is that these respondents may actually share their research data in the future due to the incentive.

The table below shows the results related to the respondents who previously indicated that they did not share research data in the past 5 years and are encouraged by the variables to share research data. Interestingly, for the variable 'low quality data', 14 respondents who previously did not share research data indicated that they

are encouraged when the data is of low quality to share it openly. The number of respondents who are stimulated by this was expected to be lower. It is not clear why these respondents do become encouraged. One possibility is the method of questioning in the questionnaire. Previous analyses have shown that the variables 'requirements by research field' and 'trust factors within data repositories' have a small probability of influence working out for respondents to share research data openly. This should also be considered in this analysis. The results show that most respondents who have not previously shared survey data are stimulated by the variables 'connection with others' and 'social norm'. In addition, the variable 'experience and skills' appears to contain many respondents who previously did not share survey data but are now encouraged to do so. This is consistent with the logistic regression in the previous section, which indicated that this variable has a positive effect on respondents' open sharing of research data.

Table 24 Number of stimulated respondents who previously did not share

Model	
Predictors	Stimulated respondents who previously did not share research data
Requirements by third parties	13 (N = 30)
Requirements by research field	14 (N = 30)
Laws and regulations	11 (N = 29)
Facilitating conditions within data repositories	8 (N = 28)
Trust factors within data repositories	13 (N = 27)
Connection with others	16 (N =27)
Consequences by sharing	11 (N = 27)
Social norm	16 (N =27)
Effort	10 (N = 27)
Use and support of data repositories	9 (N = 27)
Experience and skills	15 (N =27)
Data characteristics	9 (N =26)
Low quality data	14 (N =26)

4.3 Descriptive analyses for the barriers of sharing research data

For the third sub-question, a descriptive analysis is conducted because only the respondents who have not shared research data in the past 5 years were asked what their most common barrier is that causes them not to share research data. This descriptive analysis was conducted using SPSS. Only the results of the group of respondents not sharing research data are reflected in the output. It is important to analyze this group of respondents because only this group indicated any barriers to sharing research data. By analyzing, the results are based on only the group of respondents who did not share the research data, and not the entire group of respondents. This way it is possible to calculate how the percentages per barrier compare to the percentages of other barriers. Of the 56 respondents who completed all questions, 33 respondents did not openly share research data in the past 5 years. Therefore, the basis for this analysis is these 33 respondents. These 33 respondents identified the barrier or barriers that prevent them from sharing their research data. The table below shows which barriers are perceived as barriers by respondents who have not shared research data in the past 5 years. The frequency found for each barrier in the table is the number of respondents who indicated

that they see this as a barrier. Due to the fact that multiple responses were possible, the frequency across the table is greater than the value 33. The barrier most frequently chosen by the respondents was that the respondents had no experience in sharing the research data openly. This leads to 14 respondents. Other barriers that greatly influence not sharing the research data are that the organization involved does not want the researcher to share the research data and the research data is not shared because of sensitive information, such as privacy. For future research, it is recommended that we also ask respondents who did share research data in the past 5 years about their barriers to sharing research data openly.

Table 25 Descriptive results for the barriers

Not sharing data (N=33)	Frequency	Percentage
The data are too (privacy) sensitive	10	30.3%
One of the organizations I work with does not want me to share my data openly	11	33.3%
I don't know where to share my open research data	6	18.2%
I have no confidence in current data repositories	5	15.2%
I fear that my data would be criticized	2	6.1%
My organization does not require me to openly share my data	7	21.2%
Openly sharing data is very time-consuming	6	18.2%
I have no experience sharing open research data	14	42.4%
Other	4	12.2%

These barriers correspond to the independent variables that influence or do not influence respondents to openly share research data. For example, it can be seen from the logistic regression that the variable 'experience and skills' has a positive effect on sharing research data openly. In other words, if the researcher has the experience or skills to share research data, it may contribute to the faster sharing of research data. This also applies to the barrier 'One of the organizations I work with does not want me to share my data openly', which relates to the variable 'requirements by third parties'. This variable also has a positive effect on sharing research data openly. However, this can become a barrier if the third party, or organization indicates that they do not want research data to be shared. Similarly, this also applies to the variable 'laws and regulations', which again ties in with the barrier 'The data are too (privacy) sensitive'. This variable has a negative effect on openly sharing research data. In other words, this is a barrier to sharing research data.

Chapter 5: Discussion and recommendation

This section describes the discussion and the recommendation. This chapter described the results and it also describes the discussion points that should be considered when interpreting the results. These discussion points are specifically mentioned so that the reader will take these aspects into account and interpret the study based on these points. In a new study, these discussion points could possibly be considered. In addition, recommendations are given. In the recommendations, possible measures are given and what measures are needed to implement these measures. Several stakeholders were highlighted and several recommendations were developed for these stakeholders.

5.1 Discussion of results

For this study, a questionnaire was created and used to gather more information about the researchers' motivations for sharing or not sharing research data. This questionnaire was distributed to academic researchers working at Dutch universities specializing in the transportation discipline.

The figure below shows the modified conceptual model. This model outlines the variables for each extrinsic factor along with the effects the variables have on the dependent variable. The red line indicates a negative effect and the green line indicates a positive effect. These effects are based on the results from the logistic regression, which can be found in section 4.2.2.

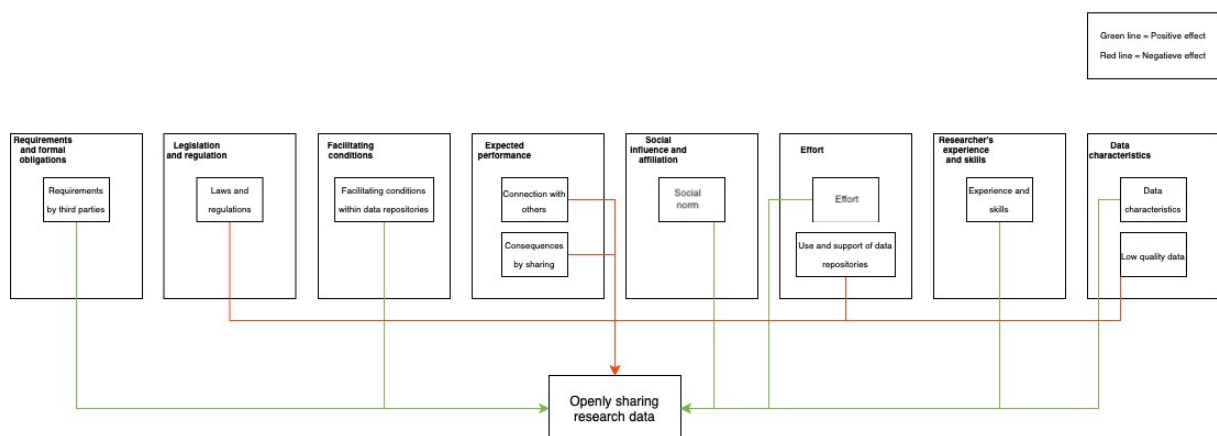


Figure 6 Conceptual model with existing effects

The results of the logistic regression show that almost all extrinsic factors that influence researchers regarding whether or not to openly share research data from other disciplines also influence researchers from the transportation discipline. This also has to do with the method of asking the questions in the questionnaire. The extrinsic factors were written out using statements where respondents could indicate whether they were stimulated or not by these statements regarding the extrinsic factors. Thus, the questionnaire did not specifically ask what extrinsic factors directly affect respondents. Thus, there is a plausible chance that for each extrinsic factor, respondents could agree with at least one statement and thus, partially, be influenced by the extrinsic factor. The influence of almost all extrinsic factors on researchers in the transportation discipline was expected because these factors were found in previous research based on 32 studies where these factors were discussed (Zuiderwijk et al., 2020). However, the analyses show that within some extrinsic factors there are more associations with sharing or not sharing research data than other extrinsic factors. The statements, which respondents were asked by extrinsic factor, were divided into a number of independent variables. These specified independent variables can be found in section 4.1. These variables were tested in the analyses. These analyses show that a number of variables have a positive effect on research data sharing, a number of variables have a negative effect on research data sharing, and a number of variables have no effect on research data sharing. The variables that positively influenced the dependent variable are 'requirements by third parties', 'facilitating conditions within data repositories', 'social norm', 'effort', 'experience and skills' and 'data

characteristics'. Where the variable 'facilitating conditions within data repositories' affects the dependent variable most strongly. The variables that negatively influences the dependent variable are 'laws and regulations', 'connection with others', 'consequences by sharing', 'use and support of data repositories' and 'low quality data'. Where the variable 'connection with others' negatively affects the variable the strongest. The two variables that appear to contribute no effect to the dependent variable are 'requirements by research field' and 'trust factors within data repositories'. The two variables weakly correlated with the dependent variable and thus have a weak relationship with the dependent variable. In addition, the p-value obtained from the logistic regression model was relatively high.

This study assumed that if the researchers did or did not share research data due to certain influencing factors in the past, they will or will not do so in the future. This assumption can be changed by adopting or implementing one of the recommendations mentioned later. These recommendations are listed in the next section.

The variable 'connection with others' has a strong negative influence on openly sharing research data. This was not initially expected. These results show that when data ends up with others or when collaborating with other researchers it does not lead researchers to openly share their research data. The same influence applies to the other variable within the extrinsic factor 'expected performance'. This variable 'consequences by sharing' also has a negative influence on the open sharing of research data. This indicates that when both positive and negative consequences are received when the research data is or is not shared, it still results in the research data not being shared. The variable 'use and support of data repositories' also negatively affects open sharing of research data. This means that when the researchers are influenced by this variable it does not lead to the open sharing of the research data. The variable 'low quality data' was expected to have a negative impact on open sharing of research data. When the researcher feels that the data is of low quality, that person is less likely to share the research data. The variable 'laws and regulations' that is associated with the extrinsic factor 'legislation and regulation' was expected to have a negative impact on open sharing of research data. A possible reason for the outcome of this study is that 'legislation and regulation' are aspects that researchers must adhere to, otherwise there are consequences. These aspects are established from higher up and apply to all people within the jurisdiction, such as the privacy aspect. These consequences apply more heavily than, for example, the consequences of the extrinsic factor 'requirements and formal obligations' because these are agreements made with the client. Part of these agreements may include asking researchers to share research data precisely. This leads to this extrinsic factor 'requirements and formal obligations', with its associated variable 'requirements by third parties' having a positive influence on the open sharing of the research data. Another extrinsic factor that has a positive effect with the associated variable 'facilitating conditions within data repositories' on sharing research data is 'facilitating conditions'. This extrinsic factor focuses on making it easier for researchers to share research data. The easier it is made for researchers, the faster the research data will be shared should other aspects, such as legislation and regulation, allow it. Another variable that has a strong influence on open sharing of research data is 'experience and skills'. This means that when researchers have the experience and skills to share research data, they do so more often. The remaining variables with a positive influence on open sharing of research data are as expected. For a detailed summary of these results, please refer to section 4.2.2.

Only those respondents who indicated that they had not shared research data in the past 5 years answered the question regarding barriers. Due to the lack of a group of respondents who gave a different answer to the dependent variable, it is not possible to conduct a logistic regression. Indeed, the effect between the barriers and the dependent variable cannot be analyzed because only respondents who did not share the research data in the past 5 years indicated the barriers. Therefore, only descriptive analysis was conducted. This descriptive analysis shows that mainly barriers related to experience, organizations involved, and laws and regulations prevent the sharing of research data. Thus, it appears that if the respondents have no experience in sharing the research data openly, they are not likely to do so in the future. In addition, it appears that when respondents

cooperate with a concerned organization, whose organization does not want the data to be shared, the researchers do not do so either. Another barrier is that the research data contains information that is too sensitive and especially in terms of privacy. The applicable laws and regulations do not allow the researchers to share the research data. The comprehensive results of the listed barriers and the other barriers can be found in the section 4.3.

5.2 Recommendation for the stakeholders

To date, research data have not been widely shared. This is reflected in the results of this study. However, it is important to share research data openly because it contributes to insightful science (Kowalczyk & Shankar, 2011). This ensures, for example, that studies are not completely duplicated because the data are already available. This saves in cost and time. In addition, sharing research data contributes to educational use, stimulation of citizen-science, and evidence-based advocacy (Kowalczyk & Shankar, 2011). **The objective to solve of this study is to identify what the extrinsic factors of researchers in the transportation discipline are and whether these extrinsic factors influence researchers in sharing research data.** To ensure that academic researchers share their research data more often in the future, measures are needed. These measures are given in this section. A couple of measures can be drawn up based on the barriers to sharing research data and the reasons for sharing research data. Several measures are outlined that can be implemented by different stakeholders.

This research has shown that the biggest barrier is when researchers have not the experience and skills to share research data, this often prevents the open sharing of research data. **One solution to this is to provide training or courses for researchers to become more experienced in sharing research data openly.** This training or courses can be offered by different parties. For example, the data repository, which provides training on how to use the particular data repository. Or an organization that provides a course to member researchers on how the organization wants research data to be shared. In addition, outside parties can also offer training or courses to openly share research data in a simple but proper way. Researchers interested in this can take these training or courses.

When researchers are not experienced in sharing research data, they are also less likely to do so. **By adding a support service to the data repository, the researchers can be helped with all their questions about sharing the research data** (Kassen, 2018; Zuiderwijk, Janssen, et al., 2014).). For example, this support team can help with the actual sharing of the research data, but they can also check the research data to see if it is suitable for open sharing. The members from this support team are then a kind of data stewards.

The second biggest barrier, not to share research data, is that researchers do not share research data because the organization they are affiliated with does not want them to. This can be solved by making sure the organization is more open to this. **The openness of the organization can be realized by establishing guidelines from within the organization so that all research data that is shared is shared in the same way and with the same data characteristics.** For example, the organization can state that only the processed research data, without any personal data, will be shared. For this, the organization can develop a software that ensures that all data, which the organization does not want to share, is filtered out and only the shareable research data remains.

It may also be that the organization to which the researcher is affiliated does not prohibit the researcher from sharing research data but does not encourage it either. When there is more encouragement from the organizations to share the research data, this will also happen more often (Kassen, 2018). **The organizations can act as promoters here.** So, it is also important that both researchers and organizations become aware that sharing research data is important. This is important for both follow-up research and new innovations as well as help for other researchers. It provides important aspects for good science, such as transparency, repetition, and progression (Huettmann, 2015). **This awareness can be created by campaigning about this within the**

organizations or platforms where these academic researchers are active (Kassen, 2018). This campaign can be conducted by, for example, the open data policy makers.

Another important barrier is the privacy aspect of the survey data (Mahajan et al., 2021). It is very important not to share personal data of respondents. Therefore, it deters researchers from sharing research data when they are dealing with personal data. Nevertheless, research data can be shared as long as all personal data is not visible. **One solution to this is to develop a software that removes or shields all personal data from the research data.** This way the researchers do not have to go through much effort to filter out the personal data and the research data can still be shared. This software can be developed by a stand-alone company or by the organization the researcher works with or the data repository where the researcher can share the research data. This measure is consistent with measures from previous studies. However, it appears that these measures have not yet been widely implemented because of the possibility of de-anonymizing the research data (Ali-Eldin et al., 2018). This ensures that there is a possibility of identifying personal information.

The other barriers to sharing research data are mainly related to data repositories. A number of respondents indicated that they do not know where to share their research data. **A solution to this might be for the developers of the data repository to launch some kind of 'marketing' campaign in places where most academic researchers are located.** This way, the data repository becomes known among academic researchers. Another possibility is to start a collaboration with an organization to which academic researchers are affiliated (Kassen, 2018). This way, academic researchers will also know more quickly where to share the research data.

Respondents indicated that when the data repository is easier to use and takes less time to share research data, they are more likely to do so. **When data repository developers focus more on the ease of sharing research data and the time it takes to share research data, it increases the likelihood that researchers will share their research data.**

There are also some respondents who indicated that they do not trust the data repository. The analysis shows that the respondents do trust a data repository realized by one of the organizations they know or are affiliated with (Morey et al., 2015). **Developers of the data repositories may start collaborations with large organizations with many academic researchers, such as universities.** Another aspect that creates more trust is the storage of research data. When data repositories guarantee that the research data is securely stored and available to others for years to come, this also generates trust. **The developers of the data repository should develop the platform so that there are no data leaks, and the data is available to the users.** A data repository can provide the platform to its users in several ways. Either it is completely open to all, or users must sign up for this platform, paid or unpaid. There will also be various forms of security attached to this. **Another aspect that creates trust is providing a communication tool from the data repository.** This communication tool allows data providers and users to communicate among themselves. For example, uncertainties in data can be explained, or collaborations can be started for further research. In addition, platform users can also communicate with the staff of the data repositories.

Research shows that respondents do not share their survey data when it is of lower quality (Schäfer et al., 2011). However, this is not always easy to assess, which is why researchers are more likely to not share their research data, rather than share it. **One solution to this is to build a quality metric into the data repository.** This quality measure will examine the quality of the research data and whether it meets the requirements of the data repository. For this, developers can develop a software that acts as a quality measure, or they can develop a software that works as a review system between multiple academic researchers in the same research field. After the research data meets the quality, it can be published. This way it is also clear to the users that they are viewing a quality research data.

Chapter 6: Conclusion

6.1 Conclusion

This chapter provides the conclusion and is aligned with the objective of the study. **The objective to solve of this study is to identify what the extrinsic factors of researchers in the transportation discipline are and whether these extrinsic factors influence researchers in sharing research data.** The research questions mentioned earlier are answered. The research questions are answered based on the observations of the results of the systematic literature review and the data analyses. Before answering the research question, it will be repeated. The research question this research answers is; *'To what extent do extrinsic factors influence openly sharing research data by academic researchers in the transportation discipline?'*. A quantitative study was conducted to answer the main question and the sub-questions.

6.1.1 Conclusion first sub-question

The first sub-question is as follows, *'What extrinsic factors influence academic researchers in openly sharing research data?'*. This first sub-question was answered using a systematic literature review. By using a systematic literature review, a structured search was conducted using keywords to answer this first sub-question. This search revealed that there are a number of factors that influence the researcher in openly sharing research data. These factors are; the researcher's background, requirements and formal obligations, personal drivers and intrinsic motivations, facilitating conditions, trust, expected performance, social influence and affiliation, effort, the researchers' experience and skills and legislation and regulation. However, these factors still include some intrinsic factors. From the systematic literature review, it became clear which factors are intrinsic and these were not further included in this study, as the focus is only on the extrinsic factors. So, the answer to the first sub-question is; the extrinsic factors that influence researchers to share research data are researcher's background, requirements and formal obligations, facilitating conditions, trust, expected performance, social influence and affiliation, effort, researcher's experience and skills, legislation and regulation and data characteristics.

6.1.2 Conclusion second sub-question

The second sub-question, *'What extrinsic factors influence researchers in the transportation discipline to share research data openly?'* was answered using a quantitative study. This quantitative research was conducted by first creating a questionnaire and then analyzing this data in the software IBM SPSS version 29. By preparing several statements for each extrinsic factor in which the respondents could indicate whether or not they agreed with it, it became clear whether the respondents, who all specialized in a research field within the transportation discipline, were influenced by the extrinsic factors to share the research data. The statements asked of respondents by extrinsic factor were converted to grouped dependent variables. Thus, the statements overlapping with each other were merged into one variable. The analysis shows that not all of independent variables that influence researchers, in general, to share research data also influence researchers in the transportation discipline. Thus, researchers from the transportation discipline are influenced by independent variables 'requirements by third parties', 'laws and regulations', 'facilitating conditions within data repositories', 'connection with others', 'consequences by sharing', 'social norm', 'effort', 'use and support of data repositories', 'experience and skills' and 'data characteristics'. These variables belong to the extrinsic factors 'requirements and formal obligations', 'legislation and regulation', 'facilitating conditions', 'expected performance', 'effort', 'social influence and affiliation', 'researcher's experience and skills' and 'data characteristics'. And so it can be concluded from this research that these extrinsic factors potentially influence academic researchers in the transportation discipline. This means that the extrinsic factor of 'trust' does not affect researchers in the transportation discipline in this research. This was found because from the tests done, none of the variables of this extrinsic factor had a correlation with the dependent variable 'sharing or not sharing research data in the past 5 years'.

6.1.3 Conclusion third sub-question

The last sub-question is as follows, *'What are the most common barriers to academic researchers in the transportation discipline to share research data openly?'*. In the questionnaire, respondents who have not shared research data in the past 5 years were asked, what are the barriers to them sharing research data. Descriptive analysis was used to analyze what barriers influence researchers not to share research data. The biggest barrier for the respondents is the lack of experience in sharing the research data openly. Respondents indicated that if they have not previously experienced openly sharing research data, they will not do so in the future. Furthermore, respondents indicated that they often do not share their research data because of agreements made with the organizations they work with. If organizations indicate that they want to keep the research data to themselves, researchers will not be permitted to share this research data. Many researchers see this as a barrier to sharing research data. Another major barrier is the sensitive information issue the data faces. Mainly, the rules regarding the privacy of the respondents keep the researchers from openly sharing the research data. Therefore, this is a barrier for which few measures can be devised, as privacy cannot be violated. Such as developing software that ensures all personal data is removed or invisible when openly sharing research data. Another measure is a development from the platform developers that provides no data leaks that can occur within the data repository.

6.1.4 Conclusion research question

The research question this research answers is; *'To what extent do extrinsic factors influence openly sharing research data by academic researchers in the transportation discipline?'*. The extent of the influence of extrinsic factors on sharing the research data is indicated by the odds ratio of the relationships with the dependent variable that resulted from the data analyses. Thus, there has been examined how many variables of the extrinsic factors have a possible relationship with the dependent variable 'sharing or not sharing research data in the past 5 years'. The logistic regression analysis shows that the variable 'facilitating conditions within data repositories', 'requirements by third parties' and 'experience and skills' have the strongest positive influence on sharing the research data over the past 5 years. In addition, the variables 'connection with others', 'low quality data' and 'laws and regulations' have the strongest negative influence on sharing research data over the past 5 years. The odds ratio represents the probability of the dependent variable by expressing the effect on the dependent variable as a percentage. All values above the value 1 for odds ratio indicate positive influence, with the higher the stronger the influence. All value below value 0 indicate a negative influence on the dependent variable. These variables belong to the extrinsic factors 'requirements and formal obligation', 'legislation and regulation', 'facilitating conditions', 'researcher's experience and skills', 'expected performance' and 'data characteristics' have the most variables with a likely probability with the dependent variable and therefore these extrinsic factors influence researchers the most. For the extrinsic factor 'data characteristics', it should still be indicated that this extrinsic factor consists of multiple variables and this factor affects the dependent variable both positively and negatively. The relationship between the independent variable and the dependent variable indicates that the researcher is influenced by the extrinsic factor to which this variable belongs. After the extrinsic factors already mentioned comes the extrinsic factor 'social influence and affiliation' which is next with the strongest influence on the dependent variable. Next comes the extrinsic factor 'effort' which both positively and negatively influences the extrinsic factor because this extrinsic factor is made up of multiple variables. The extrinsic factors that are most likely to influence researchers in openly sharing research data should be taken into account when encouraging more researchers to share research data. The recommendation provides measures to help encourage researchers to share more research data.

6.2 Scientific and theoretical contributions

This study complements previously conducted studies related to research data sharing (Zuiderwijk et al., 2020; Tedersoo et al., 2021; Tenopir et al., 2011; Zuiderwijk & Spiers, 2019). Before this study was conducted, there was none to a little focus on academic researchers from the transportation discipline and how they are influenced to share research data openly or not (Weerakkody et al., 2017). This focus is important because more focus leads to greater operational efficiency within the transportation discipline, such as reducing

congestion time or riding times (Weerakkody et al., 2017). The basis of this study is the research of Zuiderwijk et al. (2020), who listed the factors that may influence researchers in sharing research data openly. These factors are; the researcher's background, requirements and formal obligations, personal drivers and intrinsic motivations, facilitating conditions, trust, expected performance, social influence and affiliation, effort, the researchers' experience and skills and legislation and regulation. However, this study focused on researchers in the transportation discipline and on extrinsic factors. So the extrinsic factor left are; researcher's background, requirements and formal obligations, facilitating conditions, trust, expected performance, social influence and affiliation, effort, researcher's experience and skills, legislation and regulation and data characteristics. With the results from this study, it is possible to consider possible measures to ensure that researchers share their research data openly more often. Examples of these measures include; offering training by the data repository or the organization the researcher is affiliated with, adding a support service by the data repository, developing campaigns to make the open sharing of research data more attractive, adding software that removes or shields personal data in the research data and ensures that this data is not shared, and continuing to develop the data repositories that simplifies the use of these platforms.

This research contributes to open data research in general, as more has become clear about the open data aspect within the transportation discipline. This study shows that research data is not yet widely shared by researchers in the transportation sector. This is mainly due to legal, institutional, and technical factors (Hossain et al., 2015). Obviously, these factors are also included as extrinsic factors in this study. Different statements have been made based on different laws and regulations, different requirements from third parties and different technical aspects that the researchers face when openly sharing research data. Throughout this study, the extrinsic factors that influence researchers in the transportation discipline in sharing research data are examined based on several statements made in the questionnaire. Through these statements, the reasons of researchers to share research data openly or not have become clear. Based on these reasons, there are several ways to encourage researchers to share research data more often. More frequent open sharing of research data potentially leads to transparency, repetition, and progression and therefore to new innovations. These aspects lead to testing the expected performance of a study. Respondents were therefore asked if they are encouraged to share research data when the research data is useful to other researchers. These are important aspects within science (Huettmann, 2015).

Several studies have already examined the factors that influence researchers in openly sharing research data. This has indicated that these factors may vary by research discipline (Tedersoo et al., 2021; Tenopir et al., 2011; Zuiderwijk & Spiers, 2019; Zuiderwijk et al., 2020). This research continues where these studies have gone less in-depth regarding disciplines. This research focuses only on the transportation discipline. Three of the previously mentioned studies do not focus on a specific discipline but focus on data sharing in general. The study by Zuiderwijk & Spiers (2019) does focus on a discipline, namely astrophysics. Türk's (2022) study examined open research data in the discipline of epidemiology. However, this does not align with the transportation discipline. Though, this reference does align closely with this current study, as that study also looks for the factors that influence researchers on openly sharing research data.

6.3 Societal relevance and contributions

Multiple stakeholders are associated with this study (Kassen, 2018). The recommendations made can be implemented by different stakeholders. This implementation has increased the chances of openly sharing research data, leading to more research data being available to the public. The benefit to the public of the availability of research data is greater transparency. For example, studies need not be conducted frequently because the research data is already available (Kowalczyk & Shankar, 2011). This saves time and money. In addition, public trust is created by openly sharing the research data (Morey et al., 2015). This allows the public to compare data and thus be able to judge the research data for themselves. The largest group of stakeholders are the academic researchers around whom the research revolves. This group can decide whether or not to share their research data openly. Another group of stakeholders are the organizations these researchers are

associated with (Kassen, 2018). One of the most important group of stakeholders are these private organizations (Kassen, 2018). These organizations can make regulations that require or prohibit the researchers from openly sharing the research data. In addition to these organizations, there are also funding parties. These funding parties help make the research possible by providing financial support. These parties can also make rules regarding the sharing of research data. Other stakeholders are the open data policy makers at the international and national level (Hossain et al., 2015). Researchers must follow the policies that policymakers set, such as certain laws and regulations. The last stakeholders associated with this research are the developers of the data repositories (Zuiderwijk, Janssen, et al., 2014). They try to develop an orderly data repository based on the laws and regulations and the wishes of the researchers that is used as often as possible by the researchers.

Researchers, private organizations and funding parties can use this study to raise awareness of the contribution of openly sharing research data. These organizations themselves can actively encourage researchers to share research data, for example by funding. The importance of this is that more research data will be shared openly within the transportation discipline. Therefore, these private organizations can really count as promoters of the open data phenomenon (Kassen, 2018). Policy makers can use the insights from this research in formulating policies regarding open data sharing. For example, the research shows that mainly privacy-related regulations are a major restraint on open sharing of research data. Perhaps without violating privacy, policy makers can make adjustments to these regulations. In the face of drafting laws and regulations, policymakers are also promoters (Kassen, 2018). When they draft laws and regulations that are easy for researchers to combine with respect to sharing the research data, this will also be more common. The last group of stakeholders, the developers of data repositories, can do the most with the insights from this study. They can use this research in developing the data repository (Zuiderwijk, Janssen, et al., 2014). If, the respondents have indicated which aspects they would like to see reflected or not in a data repository. In addition, platform developers can also ensure that researchers and other stakeholders can interact with each other by building a communication system into the platform (Kassen, 2018). The research shows that if these aspects are present in the data repository, the research data will be shared more often.

6.4 Limitations

However, some limitations that occurred in this study must be taken into account.

The first limitation is that a limitation was created by the distribution of the questionnaire. Due to the wide distribution of the questionnaire, it was also completed by respondents outside the Netherlands. Since about half of the respondents are located in the Netherlands and the other half in the rest of the world, it is difficult to conduct geographic analyses regarding the open sharing of survey data. There are too few Dutch respondents to conduct reliable analyses regarding the geographic preferences of academic researchers and the open sharing of research data. This also applies to academic researchers outside the Netherlands.

The second limitation is that several respondents dropped out during the questionnaire. Especially for the first question, there was a decrease in respondents. One reason could be that respondents found out after the first few questions that they were not the target group. Or another reason could be that they found the questionnaire too long.

The third limitation is that it is difficult to draw reliable conclusions from the survey areas within the transportation discipline. The number of respondents per survey area is not the same, so it is difficult to draw a reliable conclusion.

Another limitation is it should be made clear that the number of respondents is not sufficient to accurately represent the population. Because some respondents did not complete the entire questionnaire and not all respondents were located in the Netherlands, the sample was reduced. For a better picture, respondents who were not located in the Netherlands were included in the analyses.

The limitations regarding questionnaire respondents were resolved by conducting the analyses anyway. In addition, consideration was given to exactly which analyses could be performed, for example, principal component analysis instead of factor analysis. This principal components analysis works better with smaller groups of respondents than the factor analysis.

As a final limitation, it is important to reiterate that the Null Hypothesis Significance Testing theory was not strictly adopted. The corresponding p-value on which this theory is based was considered a continuous value in this study (McShane et al., 2019). This means that when the p-value is greater than the value 0.05, it does not directly imply that there is no relationship between the variables. In this study, it is assumed that the lower the p-value is, the stronger the evidence is for a relationship between the variables (McShane et al., 2019). Not adopting a strict Null Hypothesis Significance Testing theory is a limitation. However, this limitation is supported by a theory. To reduce the impact of the limitation, this theory was used throughout the study.

6.5 Avenues for future research

For future research, it is recommended that a lot of time be used to find respondents. This research, because of the limited time, had 3.5 weeks to find respondents. Due to the dropout of respondents, the sample size was found to be greatly reduced. If a future study has more time, it is recommended to use a lot of time for this purpose.

If more time is available for a similar study, more time can be used in recruiting Dutch academic researchers, so that the sample consists only of researchers who have settled in the Netherlands. When public dissemination methods such as social media or mailing lists are used, respondents from other countries may also respond. Should this not be the intention for a survey, it is more convenient to write to or visit respondents individually. However, this is more time-consuming, and there is a possibility that the researcher will not have as much reach as when the questionnaire is distributed through social media or mailing lists. To obtain a large sample size, it is recommended not to focus on Dutch academic researchers in the transportation discipline, but on academic researchers in the transportation discipline worldwide.

In addition, this extra time can be used to encourage respondents to complete the questionnaire. A researcher can do this by visiting academic researchers in person. This ensures that more respondents completed the questionnaire and that more respondents completed the questionnaire in full.

This study focused on academic researchers in the transportation discipline. However, respondents were asked in which specific research field they are active. Due to the small sample size and no good distribution between these research fields, this data does not say much to conclude. This data was only used to find out if the respondents are in the transportation discipline. For future research, distinctions can also be made within the transportation discipline by targeting different research fields. It is recommended that an even sample size be taken for each research field. In this way, conclusions can be drawn regarding the research fields within the transportation discipline.

References

Academische carrière van wetenschappers | Rathenau Instituut.

(z.d.). <https://www.rathenau.nl/nl/wetenschap-cijfers/wetenschappers/personeel-aan-de-universiteiten-en-umcs/academische-carriere-van>

Ali-Eldin, A., Zuiderwijk, A., & Janssen, M. (2018). A Privacy Risk Assessment Model for Open Data. *Lecture Notes in Business Information Processing*, 186–201. https://doi.org/10.1007/978-3-319-78428-1_10

Amaresan, S. (2021, 23 november). *How to Analyze Survey Results Like a Data Pro*. <https://blog.hubspot.com/service/survey-results>

Baumeister, R. F., & Leary, M. R. (1997). Writing Narrative Literature Reviews. *Review of General Psychology*, 1(3), 311–320. <https://doi.org/10.1037/1089-2680.1.3.311>

Beer, P. D. (2022). Kwaliteit van combinatiebanen. *Tijdschrift voor HRM*, 25(2), 43–58. <https://doi.org/10.5117/thrm2022.2.cone>

Bondi, M., & Scott, M. (2010). *Keyness in Texts*. Van Haren Publishing. <https://books.google.nl/books?hl=nl&lr=&id=WUczAAAQBAJ&oi=fnd&pg=PA21&dq=key+words+research+&ots=okEdA7y35t&sig=XkCjDRN9vXfOmAS-8rSqGSOnc7w#v=onepage&q=key%20words%20research&f=false>

Campbell, J. (2015). Access to Scientific Data in the 21st Century: Rationale and Illustrative Usage Rights Review. *Data Science Journal*, 13(0), 203–230. <https://doi.org/10.2481/dsj.14-043>

Ceci, S.J. (1988). Scientists' attitudes toward data sharing. *Science, Technology, & Human Values*, 13(1/2):45–52. pmid:25309997. <https://journals.sagepub.com/doi/abs/10.1177/0162243988013001-206>

Charalabidis, Y., Zuiderwijk, A. & Janssen, M. (2012, september). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4), 258–268. <https://doi.org/10.1080/10580530.2012.716740>

Corti, L., Eynden, V. V. D., Bishop, L. & Woollard, M. (2019). *Managing and Sharing Research Data: A Guide to Good Practice* (Second). SAGE Publications Ltd.

da Costa, M.P. & Leite, .F.C.L. (2019) Factors influencing research data communication on Zika virus: a grounded theory. *Journal of Documentation*, 75(5):910–26. https://www.emerald.com/insight/content/doi/10.1108/JD-05-2018-0071/full/html?casa_token=WUBbDYoA0XEAAAAA:a7lxx8G750HfdaFnN3U-Zb5-Sw6iRApI0EN2Us7ptlgupr-M1FxTB-h4OIIIGOFcuFgi-KLLfpwdK__YJnujZQsynPlxYYdeQAUOe3ssZdPvPCfBjJk

Datadone. (2022, 28 januari). *Data processing: van ruwe data, tot hapklare informatie*. <https://datadone.nl/blog-data-processing/>

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. https://www.sciencedirect.com/science/article/pii/S1574954112000222?casa_token=CY3AENipgbQAAAAA:WCFTTddkJerBvICfX-TNnyY4kqNByXWocRluRLYfufNMqhfOlkpDGPxhdmHg6z1z1UpowmivXA

European Commission. (z.d.). *Transportation themes. Mobility and Transportation*. https://transportation.ec.europa.eu/transportation-themes_en

Extrinsiek - 4 definities - Encyclo. (z.d.). Geraadpleegd op 30 oktober 2022, van <https://www.encyclo.nl/begrip/extrinsiek>

FAQ: What are the differences between one-tailed and two-tailed tests? (z.d.). <https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faq-what-are-the-differences-between-one-tailed-and-two-tailed-tests/>

Fecher, B., Friesike, S. & Hebing, M. (2015). What drives academic data sharing? *PloS one*, 10(2):e0118053. pmid:25714752. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118053>

Feger, S., Pertiwi, C. & Bonaiuti, E. (2022). Research Data Management Commitment Drivers: An Analysis of Practices, Training, Policies, Infrastructure, and Motivation in Global Agricultural Science. In *Agricultural Research Knowledge*. <https://repo.mel.cgiar.org/handle/20.500.11766/67317>

Fields of Research (ANZSRC 2020): 3509 Transportation, Logistics and Supply Chains in Datasets - Dimensions. (z.d.). https://app.dimensions.ai/discover/data_set?and_facet_for=80081

Frost, J. (2022). *Choosing the Correct Type of Regression Analysis*. Statistics By Jim. <https://statisticsbyjim.com/regression/choosing-regression-analysis/>

Ganzevoort, W., van den Born, R.J., Halffman, W. & Turnhout, S. (2017). Sharing biodiversity data: citizen scientists' concerns and motivations. *Biodiversity and Conservation*, 1–17. <https://link.springer.com/article/10.1007/s10531-017-1391-z>

Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *Journal of Chiropractic Medicine*, 5(3), 101–117. [https://doi.org/10.1016/s0899-3467\(07\)60142-6](https://doi.org/10.1016/s0899-3467(07)60142-6)

Gritter, E. (2011). *Time Management Tips for Dissertation Writing*. https://www.stcloudstate.edu/elhe/_files/documents/dissertation/time-management-tips.pdf

Haeusermann, T., Greshake, B., Blasimme, A., Irdam, D., Richards, M. & Vayena, E. (2017). Open sharing of genomic data: Who does it and why? *PLoS ONE*, 12(5):e0177158. pmid:28486511. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0177158>

Hossain, M. A., Dwivedi, Y. K., & Rana, N. P. (2015). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of Organizational Computing and Electronic Commerce*, 26(1–2), 14–40. <https://doi.org/10.1080/10919392.2015.1124007>

How to perform a principal components analysis (PCA) in SPSS Statistics | Laerd Statistics. (z.d.). <https://statistics.laerd.com/spss-tutorials/principal-components-analysis-pca-using-spss-statistics.php>

Huettmann, F. (2015). On the Relevance and Moral Impediment of Digital Data Management, Data Sharing, and Public Open Access and Open Source Code in (Tropical) Research: The Rio Convention Revisited Towards Mega Science and Best Professional Research Practices. *Central American Biodiversity*, 391–417. https://doi.org/10.1007/978-1-4939-2208-6_16

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. https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23892?casa_token=ZPBctS80iUwAAAAA:jj2CFsO_Gdf0-QcKQgkLh60cL_lvLm_qMJHtvUwEiCkHyJpoSNyhPTQnPBpaU8YVfBXXrSN2SZIB-k

Kassen, M. (2018). Adopting and managing open data: Stakeholder perspectives, challenges and policy recommendations. *Aslib Journal of Information Management*, 70(5).

<https://www.emerald.com/insight/content/doi/10.1108/AJIM-11-2017-0250/full/html>

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.

https://www.sciencedirect.com/science/article/pii/S0268401215000432?casa_token=CSIBbTiUfsMAAAAA:q57bjoPYts1NFE9nid_TyfRg35FXX9mYBHjyaJlkity1KwMkeYTtlzDqPMDRDteYOyGj3ln9qQ

Kim, Y. & Yoon, A. (2017). Scientists' data reuse behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, 68(12):2709–19.

https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23892?casa_token=jSzW9PUR1gkAAAA:Mb3TJpsYKVBjInKg9YOG0t3rE2-D2n1NlnpFd9APGD9IB8eq_tjORPMt07ezvGuaocf9QGO4RfEA0FA

Kitchenham, B.A. (2004). *Procedures for Performing Systematic Reviews*. Software Engineering Group Department of Computer Science. <https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf>

Kitchin, R. (2014, 26 augustus). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (1ste editie). SAGE Publications Ltd.

Kowalczyk, S., & Shankar, K. (2011). Data sharing in the sciences. *Annual Review of Information Science and Technology*, 45(1), 247–294. <https://doi.org/10.1002/aris.2011.1440450113>

Mahajan, V., Kuehnel, N., Intzevidou, A., Cantelmo, G., Moeckel, R. & Antoniou, C. (2021, 15 september). Data to the people: a review of public and proprietary data for transportation models. *Transportation Reviews*, 42(4), 415–440. <https://doi.org/10.1080/01441647.2021.1977414>

McShane, B. B., Gal, D., Gelman, A., Robert, C. & Tackett, J. L. (2019). Abandon Statistical Significance. *The American Statistician*, 73(sup1), 235–245. <https://doi.org/10.1080/00031305.2018.1527253>

Meijer, B. (2018). Een nieuw fundament: beeld van de technieksector. *sectorplan-betatechniek*. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKewil_bC9pPv7AhWZO-wKHakSBIUQFnoECA4QAw&url=https%3A%2F%2Fwww.sectorplan-betatechniek.nl%2Fsites%2Fsectorplan_bt%2Ffiles%2Fmedia-files%2FSectorbeeld%252BTechniek.pdf&usq=AOvVaw3CDf2-5_GR5gNXPMgts5oA

Molin, E. (z.d. -a). *Logistische regressie* [Presentatieslides].

Brightspace. <https://brightspace.tudelft.nl/d2l/le/content/148558/viewContent/1193695/View>

Molin, E. (z.d. -b). *Principale componenten analyse* [Presentatieslides].

Brightspace. <https://brightspace.tudelft.nl/d2l/le/content/148558/viewContent/1177980/View>

Morey, T., Forbath, T. & Scoop, A. (2015). Customer data: Designing for transparency and trust. *Harvard Business Review*, 93(5), 96–105. https://www.academia.edu/download/49352349/CUSTOMER_DATA-DESIGNING_FOR_TRANSPARENCY_AND_TRUST-R1505H-PDF-ENG.desbloqueado.pdf

Multipele lineaire regressie. (z.d.). <https://sh-studiedata.nl/R/28-Multipele-lineaire-regressie-R.html>

Murray-Rust, P. (2008, 18 januari). Open Data in Science. *Nature Precedings*.

<https://doi.org/10.1038/npre.2008.1526.1>

One-Sample Proportions. (z.d.). <https://www.ibm.com/docs/en/spss-statistics/saas?topic=proportions-one-sample>

Osagie, E., Waqar, M., Adebayo, S., Stasiewicz, A., Porwol, L. & Ojo, A. (2017, 7 juni). Usability Evaluation of an Open Data Platform. *Proceedings of the 18th Annual International Conference on Digital Government Research*. <https://doi.org/10.1145/3085228.3085315>

People. (z.d.). TU Delft. Geraadpleegd op 15 oktober 2022, van <https://www.tudelft.nl/en/tpm/our-faculty/departments/engineering-systems-and-services/people>

Piwowar, H.A., Day, R.S. & Fridsma, D.B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2: e308(3):1–5. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308>

Piwowar, H. A. & Vision, T. J. (2013, 1 oktober). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. <https://doi.org/10.7717/peerj.175>

Publieke verkeers- en transportationdata. (2020). Talking Traffic. <https://www.talking-traffic.com/images/Documenten/DASHBOARD-NL.pdf>

Raffaghelli JE, Manca S. Is there a social life in open data? The case of open data practices in educational technology research. *Publications*. 2019;7(1):9. <https://www.mdpi.com/402596>

RDM explained. (z.d.). Universiteit Gent. <https://www.ugent.be/en/research/datamanagement/why/rdm-explained.htm>

Roopa, S. & Rani, M. (2012). Questionnaire Designing for a Survey. *Journal of Indian Orthodontic Society*, 46(4_suppl1), 273–277. <https://doi.org/10.1177/0974909820120509s>

Rother, E. T. (2007). Systematic literature review X narrative review. *Acta Paulista De Enfermagem*, 20(2). <https://doi.org/10.1590/s0103-21002007000200001>

Ryan, R. M. & Deci, E. L. (2000, januari). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1), 54–67. <https://doi.org/10.1006/ceps.1999.1020>

Sá, C. & Grieco, J. (2016). Open data for science, policy, and the public good. *Review of Policy Research*, 33(5):526–43. https://onlinelibrary.wiley.com/doi/abs/10.1111/ropr.12188?casa_token=1qFkUn5bnrUAAAAA:WnlHWxRXCgGq6pjCF5thmFA5aTdEHj8xJYYJoy-Byx5m7Lmf8U_iJe3ed2auHeOSsrfRw33dSvqz4

Schäfer, A., Pampel, H., Pfeiffenberger, H., Dallmeier-Tiessen, S., Tissari, S., Darby, R., Giaretta, K., Giaretta, D., Gitmans, K., Helin, H., Lambert, S., Mele, S., Reilly, S., Ruiz, S., Sandberg, M., Schallier, W., Schrimpf, S., Smit, E., Wilkinson, M. & Wilson, M. (2011). Baseline Report on Drivers and Barriers in Data Sharing. *Opportunities for Data Exchange (ODE)*. <https://doi.org/10.5281/zenodo.8306>

Schmidt, B., Gemeinholzer, B. & Treloar, A. (2016). Open data in global environmental research: The Belmont Forum's open data survey. *PloS one*, 11(1):e0146695. [pmid:26771577. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146695](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146695)

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia Medica*, 12–18. <https://doi.org/10.11613/bm.2014.003>

Statistiek: Chi-kwadraat voor samenhang. (z.d.). InfoNu. <https://wetenschap.infonu.nl/wiskunde/122070-statistiek-chi-kwadraat-voor-samenhang.html>

Steekproefcalculator. (2020, 11 februari). CheckMarket. Geraadpleegd op 15 oktober 2022, van <https://nl.checkmarket.com/steekproefcalculator/>

- Stevens, E. (2022, 13 oktober). *The 7 Most Useful Data Analysis Methods and Techniques*. CareerFoundry. <https://careerfoundry.com/en/blog/data-analytics/data-analysis-techniques/>
- Taguchi, N. (2018). Description and explanation of pragmatic development: Quantitative, qualitative, and mixed methods research. *System*, 75, 23–32. <https://doi.org/10.1016/j.system.2018.03.010>
- Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, L., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K. & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, 8(1). <https://doi.org/10.1038/s41597-021-00981-0>
- Tenkanen, H., Saarsalmi, P., Järv, O., Salonen, M. & Toivonen, T. (2016, 28 juli). Health research needs more comprehensive accessibility measures: integrating time and transportation modes from open data. *International Journal of Health Geographics*, 15(1). <https://doi.org/10.1186/s12942-016-0052-x>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M. & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE*, 6(6), e21101. <https://doi.org/10.1371/journal.pone.0021101>
- The Chi-Square Assumption States That Each Cell of The 2x2 Table Must Have Five Observations.* (z.d.). Accredited Professional Statistician For Hire. <https://www.scalestatistics.com/chi-square-assumption.html>
- TU Delft. (z.d.). *SEN2321 Research methods and tools final* [Presentatieslides]. Greatlibrary.com. <https://greatlibrary.nl/content//SEN2321%20Research%20methods%20and%20tools%20final.ppt>
- Türk, B. (2020). Enhancing Open Research Data Sharing and Reuse via Infrastructural and Institutional Instruments: a Case Study in Epidemiology. TU Delft.
- Universiteit Utrecht. (z.d.). *Chi-kwadraat toets*. https://www.uu.nl/sites/default/files/fsw_ms_spss_chi-kwadraat_toets.pdf
- Verhoeven, N. (2018). *Wat is onderzoek?: Praktijkboek voor methoden en technieken*. Boom Lemma. https://www.bua.nl/media/16/9789024406937_-_nel_verhoeven_-_wat_is_onderzoek_zesde_druk_-_inkijkexemplaar.pdf
- Wallis, J.C., Rolando, E. & Borgman, C.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):e67332. pmid:23935830. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0067332>
- Weerakkody, V., Kapoor, K., Balta, M. E., Irani, Z. & Dwivedi, Y. K. (2017, 11 juli). Factors influencing user acceptance of public sector big open data. *Production Planning & Control*, 28(11–12), 891–905. <https://doi.org/10.1080/09537287.2017.1336802>
- Yoon, A. & Kim, Y. (2017). Social scientists' data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories. *Library & Information Science Research*, 39(3):224–33. https://www.sciencedirect.com/science/article/pii/S0740818816302250?casa_token=5BEYwcVkixkAAAAA:Kj821BvQrD2qjkab1qBVGJ2rbmWTrSuBo3tTyNtZfXtHPsmx8KJrHq384cvEQtzvedkVYesrog
- Zee, F. van der. (2022, 6 juni). *Dummy-variabele of indicatorvariabele*. Hulp bij Onderzoek. <https://hulpbijonderzoek.nl/online-woordenboek/begrippen/dummy-variabele/>
- Zenk-Möltgen, W., Akdeniz, E., Katsanidou, A., Naßhoven, V. & Balaban, E. (2018). Factors influencing the data sharing behavior of researchers in sociology and political science. *Journal of documentation*, 74(5):1053–73. https://www.emerald.com/insight/content/doi/10.1108/JD-09-2017-0126/full/html?casa_token=-

IAAi4mEZcgAAAAA:ox0zJ8Eule1WmetE-QvkG-zaMoRYQjmFwGWXJ06BaRp0jacNQdi6W-
QPxB34CyNXUgeieK5ds5I7puRw5Uc58Viq_c_184hh106DNzgiZIFE1xE2XP0

Zimmerman, A. (2007). Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries*, 7(1–2):5–16.

<http://scholar.google.com/scholar?q=Not+by+metadata+alone%3A+the+use+of+diverse+forms+of+knowledge+to+locate+data+for+reuse+Zimmerman+2007>

Zuiderwijk, A., Helbig, N., Gil-Garcia, J. & Janssen, M. (2014, augustus). Special Issue on Innovation through Open Data - A Review of the State-of-the-Art and an Emerging Research Agenda: Guest Editors' Introduction. *Journal of theoretical and applied electronic commerce research*, 9(2), 1–2.
<https://doi.org/10.4067/s0718-18762014000200001>

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, 19(1,2), 17–33. <https://doi.org/10.3233/ip-140329>

Zuiderwijk, A., Shinde, R. & Jeng, W. (2020, 18 september). What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. *PLOS ONE*, 15(9), e0239283. <https://doi.org/10.1371/journal.pone.0239283>

Zuiderwijk, A. & Spiers, H. (2019, december). Sharing and re-using open data: A case study of motivations in astrophysics. *International Journal of Information Management*, 49, 228–241.
<https://doi.org/10.1016/j.ijinfomgt.2019.05.024>

Appendices

Appendix A. Studies about open data sharing per discipline

The table below lists literature focusing on the various disciplines.

Table 26 Studies about open data sharing per discipline

Discipline	Literature
Health disciplines	da Costa, M.P., Leite FCL. Factors influencing research data communication on Zika virus: a grounded theory. <i>Journal of Documentation</i> . 2019;75(5):910–26. https://www.emerald.com/insight/content/doi/10.1108/JD-05-2018-0071/full/html?casa_token=WUBbDYoA0XEAAAAA:a7lx8G750HfdaFnN3U-Zb5-Sw6iRApI0EN2Us7ptIgupr-M1FXTB-h40IIGOFcuFgi-KLLfpwdK_YJnujZQsynPkyYdeQAU0e3ssZdPvPCfBJjk
	Joo S, Kim S, Kim Y. An exploratory study of health scientists' data reuse behaviors: Examining attitudinal, social, and resource factors. <i>Aslib Journal of Information Management</i> . 2017;69(4):389–407. https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23892?casa_token=ZPBCT580iUwAAAAA:jj2CFsO_Gdfo-QcKqgKlh60cl_vLm_qMJHtvUwEiCkHyJpoSNyhPTQNpBpaU8YVfBxxrSN2SZIB-k
	Piwowar, H.A., Day, R.S., Fridsma DB. Sharing detailed research data is associated with increased citation rate. <i>PLoS ONE</i> . 2007;2: e308(3):1–5. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308
	Tenkanen, H., Saarsalmi, P., Järvi, O., Salonen, M. & Toivonen, T. (2016, 28 juli). Health research needs more comprehensive accessibility measures: integrating time and transportation modes from open data. <i>International Journal of Health Geographics</i> , 15(1). https://doi.org/10.1186/s12942-016-0052-x
Biology disciplines	Ganzevoort W, van den Born, R.J., Halffman W, Turnhout S. Sharing biodiversity data: citizen scientists' concerns and motivations. <i>Biodiversity and Conservation</i> . 2017:1–17. https://link.springer.com/article/10.1007/s10531-017-1391-z
	Haeusermann T, Greshake B, Blasimme A, Irdam D, Richards M, Vayena E. Open sharing of genomic data: Who does it and why? <i>PLoS ONE</i> . 2017;12(5):e0177158. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0177158
	Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B. The user's view on biodiversity data sharing—Investigating facts of acceptance and requirements to realize a sustainable use of research data <i>Ecological Informatics</i> . 2012;11:25–33. https://www.sciencedirect.com/science/article/pii/S1574954112000222?casa_token=CY3AENipgbQAAAAA:WCFTTddkIerBvICfX-TNny4kqNByXWocRiURLYufNMqhfOLkDPxhdmHg6zIz1UpowmivXA
	Feger, S., Pertiwi, C. & Bonaiuti, E. (2022). Research Data Management Commitment Drivers: An Analysis of Practices, Training, Policies, Infrastructure, and Motivation in Global Agricultural Science. In <i>Agricultural Research Knowledge</i> . https://repo.mel.cgiar.org/handle/20.500.11766/67317
Social disciplines	Kim Y, Adler M. Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. <i>International Journal of Information Management</i> . 2015;35(4):408–18. https://www.sciencedirect.com/science/article/pii/S0268401125000432?casa_token=CSIBbTUIfsMAAAAA:q57bjoPYts1NFE9nid_Tyfrg35FXX9mYBHjyaJlkity1KwMk eYTtZDqPMDRDteYOyGj3In9qQ
	Yoon A, Kim Y. Social scientists' data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories. <i>Library & Information Science Research</i> . 2017;39(3):224–33. https://www.sciencedirect.com/science/article/pii/S0740818816302250?casa_token=5BEYwcVkiXkAAAAA:Kj82IBvQrD2qjkab1qBVGJ2rbmWTrSuBo3tTyNTzFxtHPS Mx8KJrHq384cvEQZvedkVYesrog
	Zenk-Möltgen W, Akdeniz E, Katsanidou A, Naßhoven V, Balaban E. Factors influencing the data sharing behavior of researchers in sociology and political science. <i>Journal of documentation</i> . 2018;74(5):1053–73. https://www.emerald.com/insight/content/doi/10.1108/JD-09-2017-0126/full/html?casa_token=IAAi4mEZcgAAAAA:ox0zJ8Eule1WmetE-QvkG-zaMoRYQjmfWGWXJ06BaRp0jacNQdi6W-QPxB34CynXUgeieK5ds5i7puRw5Uc58Viq_c_184hh106DNzgiZIFE1xe2XP0
Scientific disciplines	Sá C, Grieco J. Open data for science, policy, and the public good. <i>Review of Policy Research</i> . 2016;33(5):526–43. https://onlinelibrary.wiley.com/doi/abs/10.1111/ropr.12188?casa_token=1qFkUn5bnrUAAAAA:WnlHWxRXCGGq6pjCF5thmFA5aTdEHj8xjYYJoy-Byx5m7Lmf8U_ije3ed2auHeOSsrRw33dSvqz4
	Raffaghelli JE, Manca S. Is there a social life in open data? The case of open data practices in educational technology research. <i>Publications</i> . 2019;7(1):9. https://www.mdpi.com/402596
	Wallis, J.C., Rolando E, Borgman CL. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. <i>PLoS ONE</i> . 2013;8(7):e67332. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0067332
Environmental disciplines	Schmidt B, Gemeinholzer B, Treloar A. Open data in global environmental research: The Belmont Forum's open data survey. <i>PLoS one</i> . 2016;11(1):e0146695. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146695
	Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. <i>PLoS one</i> . 2011;6(6):e21101. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101
	Zimmerman A. Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. <i>International Journal on Digital Libraries</i> . 2007;7(1–2):5–16. http://scholar.google.com/scholar?q=Not+by+metadata+alone%3A+the+use+of+diverse+forms+of+knowledge+to+locate+data+for+reuse+Zimmerman+2007

Astrophysics discipline	Zuiderwijk A, Spiers H. Sharing and re-using open data: A case study of motivations in astrophysics. <i>International Journal of Information Management</i> . 2019;49:228–41. https://www.sciencedirect.com/science/article/pii/S0268401218311836
Transportation discipline	Mahajan, V., Kuehnel, N., Intzeidou, A., Cantelmo, G., Moeckel, R. & Antoniou, C. (2021, 15 september). Data to the people: a review of public and proprietary data for transportation models. <i>Transportation Reviews</i> , 42(4), 415–440. https://doi.org/10.1080/01441647.2021.1977414
Various disciplines	Kim Y, Yoon A. Scientists' data reuse behaviors: A multilevel analysis. <i>Journal of the Association for Information Science and Technology</i> . 2017;68(12):2709–19. https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23892?casa_token=JszW9PUR1gkAAAAA:Mb3TJpsYKVBjInKg9YOG0t3rE2-D2n1NlnpFd9APGD9IB8eq_tj0RPMt07ezvGuaocf9QGO4RfEA0FA
	Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PloS one</i> . 2015;10(2):e0118053. pmid:25714752. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118053
	Ceci, S.J.. Scientists' attitudes toward data sharing. <i>Science, Technology, & Human Values</i> . 1988;13(1/2):45–52. pmid:25309997. https://journals.sagepub.com/doi/abs/10.1177/0162243988013001-206
	Weerakkody, V., Kapoor, K., Balta, M. E., Irani, Z. & Dwivedi, Y. K. (2017, 11 juli). Factors influencing user acceptance of public sector big open data. <i>Production Planning & Control</i> , 28(11–12), 891–905. https://doi.org/10.1080/09537287.2017.1336802

Appendix B. The sources of the extrinsic factors influencing data sharing

Table 27 The sources of the extrinsic factors influencing data sharing

Categories	Factors	Literature
Researcher's background	Culture	Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PLoS one</i> . 2015;10(2):e0118053. PMID:25714752
		da Costa MP, Leite FCL. Factors influencing research data communication on Zika virus: a grounded theory. <i>Journal of Documentation</i> . 2019;75(5):910–26
		Zuiderwijk A, Spiers H. Sharing and re-using open data: A case study of motivations in astrophysics. <i>International Journal of Information Management</i> . 2019;49:228–41.
	Country	Joo S, Kim S, Kim Y. An exploratory study of health scientists' data reuse behaviors: Examining attitudinal, social, and resource factors. <i>Aslib Journal of Information Management</i> . 2017;69(4):389–407.
		Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. <i>PLoS one</i> . 2011;6(6):e21101. PMID:21738610
		Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PLoS one</i> . 2015;10(2):e0118053. PMID:25714752
	Research field	Joo S, Kim S, Kim Y. An exploratory study of health scientists' data reuse behaviors: Examining attitudinal, social, and resource factors. <i>Aslib Journal of Information Management</i> . 2017;69(4):389–407.
		Curry RG, Crowston K, Specht A, Grant BW, Dalton ED. Attitudes and norms affecting scientists' data reuse. <i>PLoS ONE</i> . 2017;12(12):e0189288. PMID:29281658
		Kim Y, Yoon A. Scientists' data reuse behaviors: A multilevel analysis. <i>Journal of the Association for Information Science and Technology</i> . 2017;68(12):2709–19.
		Yoon A, Kim Y. Social scientists' data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories. <i>Library & Information Science Research</i> . 2017;39(3):224–33.
Requirements and formal obligations	Mandates from parties	Curry RG, Crowston K, Specht A, Grant BW, Dalton ED. Attitudes and norms affecting scientists' data reuse. <i>PLoS ONE</i> . 2017;12(12):e0189288. PMID:29281658
		Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. <i>PLoS one</i> . 2011;6(6):e21101. PMID:21738610
	Funding	Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PLoS one</i> . 2015;10(2):e0118053. PMID:25714752
		Kim Y, Adler M. Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. <i>International Journal of Information Management</i> . 2015;35(4):408–18.
		Schmidt B, Gemeinholzer B, Treloar A. Open data in global environmental research: The Belmont Forum's open data survey. <i>PLoS one</i> . 2016;11(1):e0146695. PMID:26771577
		Mooney H, Newton MP. The anatomy of a data citation: Discovery, reuse, and credit. <i>Journal of Librarianship and Scholarly Communication</i> . 2012;1(1):eP1035.
		Schmidt B, Gemeinholzer B, Treloar A. Open data in global environmental research: The Belmont Forum's open data survey. <i>PLoS one</i> . 2016;11(1):e0146695. PMID:26771577
	Compliance policies	Curry RG, Crowston K, Specht A, Grant BW, Dalton ED. Attitudes and norms affecting scientists' data reuse. <i>PLoS ONE</i> . 2017;12(12):e0189288. PMID:29281658
	Ethics codes	Harper LM, Kim Y. Attitudinal, normative, and resource factors affecting psychologists' intentions to adopt an open data badge: An empirical analysis. <i>International Journal of Information Management</i> . 2018;41:23–32.
Facilitating conditions	Technical infrastructure	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. <i>Data Science Journal</i> . 2004;3(29):135–52.
		Joo S, Kim S, Kim Y. An exploratory study of health scientists' data reuse behaviors: Examining attitudinal, social, and resource factors. <i>Aslib Journal of Information Management</i> . 2017;69(4):389–407.
		Wallis JC, Rolando E, Borgman CL. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. <i>PLOS ONE</i> . 2013;8(7):e67332. PMID:23935830
	Funding	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. <i>Data Science Journal</i> . 2004;3(29):135–52.
		Cragin MH, Palmer CL, Carlson JR, Witt M. Data sharing, small science and institutional repositories. <i>Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences</i> . 2010;368(1926):4023–38.
		da Costa MP, Leite FCL. Factors influencing research data communication on Zika virus: a grounded theory. <i>Journal of Documentation</i> . 2019;75(5):910–26.
		Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PLoS one</i> . 2015;10(2):e0118053. PMID:25714752
		Harper LM, Kim Y. Attitudinal, normative, and resource factors affecting psychologists' intentions to adopt an open data badge: An empirical analysis. <i>International Journal of Information Management</i> . 2018;41:23–32
	Information systems	da Costa MP, Leite FCL. Factors influencing research data communication on Zika virus: a grounded theory. <i>Journal of Documentation</i> . 2019;75(5):910–26.
	Data repositories	Mooney H, Newton MP. The anatomy of a data citation: Discovery, reuse, and credit. <i>Journal of Librarianship and Scholarly Communication</i> . 2012;1(1):eP1035
		Joo S, Kim S, Kim Y. An exploratory study of health scientists' data reuse behaviors: Examining attitudinal, social, and resource factors. <i>Aslib Journal of Information Management</i> . 2017;69(4):389–407.
		Harper LM, Kim Y. Attitudinal, normative, and resource factors affecting psychologists' intentions to adopt an open data badge: An empirical analysis. <i>International Journal of Information Management</i> . 2018;41:23–32.

		<p>Kim Y, Adler M. Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. <i>International Journal of Information Management</i>. 2015;35(4):408–18.</p> <p>da Costa MP, Leite FCL. Factors influencing research data communication on Zika virus: a grounded theory. <i>Journal of Documentation</i>. 2019;75(5):910–26.</p>
	Facilitation platforms	Zuiderwijk A, Spiers H. Sharing and re-using open data: A case study of motivations in astrophysics. <i>International Journal of Information Management</i> . 2019;49:228–41.
	Communication	<p>Kim Y, Yoon A. Scientists' data reuse behaviors: A multilevel analysis. <i>Journal of the Association for Information Science and Technology</i>. 2017;68(12):2709–19.</p> <p>Zuiderwijk A. Open data infrastructures: The design of an infrastructure to enhance the coordination of open data use. 's-Hertogenbosch: Uitgeverij BOXPress; 2015.</p>
	Accessibility	Campbell J. Access to scientific data in the 21st century: Rationale and illustrative usage rights review. <i>Data Science Journal</i> . 2015;13:203–30.
Trust	Transparency	Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B. The user's view on biodiversity data sharing—Investigating facts of acceptance and requirements to realize a sustainable use of research data <i>Ecological Informatics</i> . 2012;11:25–33.
	Validity	Sá C, Grieco J. Open data for science, policy, and the public good. <i>Review of Policy Research</i> . 2016;33(5):526–43
	Data management	<p>Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. <i>PLoS one</i>. 2011;6(6):e21101. pmid:21738610</p> <p>Campbell J. Access to scientific data in the 21st century: Rationale and illustrative usage rights review. <i>Data Science Journal</i>. 2015;13:203–30.</p>
	Data security	Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PLoS one</i> . 2015;10(2):e0118053. pmid:25714752
	Simplifying interpretation	<p>Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PLoS one</i>. 2015;10(2):e0118053. pmid:25714752</p> <p>Cragin MH, Palmer CL, Carlson JR, Witt M. Data sharing, small science and institutional repositories. <i>Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences</i>. 2010;368(1926):4023–38.</p> <p>Schmidt B, Gemeinholzer B, Treloar A. Open data in global environmental research: The Belmont Forum's open data survey. <i>PLoS one</i>. 2016;11(1):e0146695. pmid:26771577</p> <p>Zuiderwijk A. Open data infrastructures: The design of an infrastructure to enhance the coordination of open data use. 's-Hertogenbosch: Uitgeverij BOXPress; 2015.</p> <p>Harper LM, Kim Y. Attitudinal, normative, and resource factors affecting psychologists' intentions to adopt an open data badge: An empirical analysis. <i>International Journal of Information Management</i>. 2018;41:23–32.</p> <p>Sayogo DS, Pardo T. Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. <i>Government Information Quarterly</i>. 2013;30(1):S19–S31.</p>
Expected performance	Criticism	Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B. The user's view on biodiversity data sharing—Investigating facts of acceptance and requirements to realize a sustainable use of research data <i>Ecological Informatics</i> . 2012;11:25–33.
	Recognition	<p>Harper LM, Kim Y. Attitudinal, normative, and resource factors affecting psychologists' intentions to adopt an open data badge: An empirical analysis. <i>International Journal of Information Management</i>. 2018;41:23–32.</p> <p>Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PLoS one</i>. 2015;10(2):e0118053. pmid:25714752</p> <p>da Costa MP, Leite FCL. Factors influencing research data communication on Zika virus: a grounded theory. <i>Journal of Documentation</i>. 2019;75(5):910–26.</p> <p>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. <i>Data Science Journal</i>. 2004;3(29):135–52.</p>
	Reward systems	<p>Mooney H, Newton MP. The anatomy of a data citation: Discovery, reuse, and credit. <i>Journal of Librarianship and Scholarly Communication</i>. 2012;1(1):eP1035.</p> <p>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. <i>Data Science Journal</i>. 2004;3(29):135–52.</p>
	Data visibility	<p>Mooney H, Newton MP. The anatomy of a data citation: Discovery, reuse, and credit. <i>Journal of Librarianship and Scholarly Communication</i>. 2012;1(1):eP1035</p> <p>Joo S, Kim S, Kim Y. An exploratory study of health scientists' data reuse behaviors: Examining attitudinal, social, and resource factors. <i>Aslib Journal of Information Management</i>. 2017;69(4):389–407.</p> <p>da Costa MP, Leite FCL. Factors influencing research data communication on Zika virus: a grounded theory. <i>Journal of Documentation</i>. 2019;75(5):910–26.</p>
	Researcher's profile	Cragin MH, Palmer CL, Carlson JR, Witt M. Data sharing, small science and institutional repositories. <i>Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences</i> . 2010;368(1926):4023–38
	Collaboration	<p>Harper LM, Kim Y. Attitudinal, normative, and resource factors affecting psychologists' intentions to adopt an open data badge: An empirical analysis. <i>International Journal of Information Management</i>. 2018;41:23–32.</p> <p>Zuiderwijk A, Spiers H. Sharing and re-using open data: A case study of motivations in astrophysics. <i>International Journal of Information Management</i>. 2019;49:228–41.</p>
	Data accessibility	<p>Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B. The user's view on biodiversity data sharing—Investigating facts of acceptance and requirements to realize a sustainable use of research data <i>Ecological Informatics</i>. 2012;11:25–33.</p> <p>Schmidt B, Gemeinholzer B, Treloar A. Open data in global environmental research: The Belmont Forum's open data survey. <i>PLoS one</i>. 2016;11(1):e0146695. pmid:26771577</p>
	Funding	Mooney H, Newton MP. The anatomy of a data citation: Discovery, reuse, and credit. <i>Journal of Librarianship and Scholarly Communication</i> . 2012;1(1):eP1035.
Social influence and affiliation	Pressure	<p>Kim Y, Adler M. Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. <i>International Journal of Information Management</i>. 2015;35(4):408–18.</p> <p>Harper LM, Kim Y. Attitudinal, normative, and resource factors affecting psychologists' intentions to adopt an open data badge: An empirical analysis. <i>International Journal of Information Management</i>. 2018;41:23–32</p>

		Zenk-Möltgen W, Akdeniz E, Katsanidou A, Naßhoven V, Balaban E. Factors influencing the data sharing behavior of researchers in sociology and political science. <i>Journal of documentation</i> . 2018;74(5):1053–73.
		Curty RG, Crowston K, Specht A, Grant BW, Dalton ED. Attitudes and norms affecting scientists' data reuse. <i>PLoS ONE</i> . 2017;12(12):e0189288. PMID:29281658
Effort	Degree of use of data	Curty RG, Crowston K, Specht A, Grant BW, Dalton ED. Attitudes and norms affecting scientists' data reuse. <i>PLoS ONE</i> . 2017;12(12):e0189288. PMID:29281658 Piwowar HA, Vision TJ. Data reuse and the open data citation advantage. <i>PeerJ</i> . 2013;1:e175. PMID:24109559 Yoon A. Data reusers' trust development. <i>Journal of the Association for Information Science and Technology</i> . 2017;68(4):946–56.
	Ease of use of facilitating platforms	da Costa MP, Leite FCL. Factors influencing research data communication on Zika virus: a grounded theory. <i>Journal of Documentation</i> . 2019;75(5):910–26. Cragin MH, Palmer CL, Carlson JR, Witt M. Data sharing, small science and institutional repositories. <i>Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences</i> . 2010;368(1926):4023–38 Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. <i>PLoS one</i> . 2011;6(6):e21101. PMID:21738610 Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B. The user's view on biodiversity data sharing—Investigating facts of acceptance and requirements to realize a sustainable use of research data <i>Ecological Informatics</i> . 2012;11:25–33 Wallis JC, Rolando E, Borgman CL. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. <i>PLOS ONE</i> . 2013;8(7):e67332. PMID:23935830
	(Technical) support	Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PLoS one</i> . 2015;10(2):e0118053. PMID:25714752
	Recognition	Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B. The user's view on biodiversity data sharing—Investigating facts of acceptance and requirements to realize a sustainable use of research data <i>Ecological Informatics</i> . 2012;11:25–33
	Time investment	Wallis JC, Rolando E, Borgman CL. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. <i>PLOS ONE</i> . 2013;8(7):e67332. PMID:23935830 Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. <i>PLoS ONE</i> . 2007;2: e308(3):1–5. Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B. The user's view on biodiversity data sharing—Investigating facts of acceptance and requirements to realize a sustainable use of research data <i>Ecological Informatics</i> . 2012;11:25–33 Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PLoS one</i> . 2015;10(2):e0118053. PMID:25714752 Sayogo DS, Pardo T. Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. <i>Government Information Quarterly</i> . 2013;30(1):S19–S31.
Researcher's experience and skills	Data for learning and training purposes	Kim Y, Adler M. Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. <i>International Journal of Information Management</i> . 2015;35(4):408–18. Arza V, Fressoli M. Systematizing benefits of open science practices. <i>Information Services & Use</i> . 2017;37(4):463–74.
Legislation and regulation	Laws and regulations	Zuiderwijk A, Spiers H. Sharing and re-using open data: A case study of motivations in astrophysics. <i>International Journal of Information Management</i> . 2019;49:228–41.
	Data sharing	Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PLoS one</i> . 2015;10(2):e0118053. PMID:25714752 Mooney H, Newton MP. The anatomy of a data citation: Discovery, reuse, and credit. <i>Journal of Librarianship and Scholarly Communication</i> . 2012;1(1):eP1035. Joo S, Kim S, Kim Y. An exploratory study of health scientists' data reuse behaviors: Examining attitudinal, social, and resource factors. <i>Aslib Journal of Information Management</i> . 2017;69(4):389–407. Schmidt B, Gemeinholzer B, Treloar A. Open data in global environmental research: The Belmont Forum's open data survey. <i>PLoS one</i> . 2016;11(1):e0146695. PMID:26771577
	Data transparency	Ganzevoort W, van den Born RJ, Halfman W, Turnhout S. Sharing biodiversity data: citizen scientists' concerns and motivations. <i>Biodiversity and Conservation</i> . 2017;1–17
	National laws and regulation	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. <i>Data Science Journal</i> . 2004;3(29):135–52.
	Rules from clients	Sayogo DS, Pardo T. Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. <i>Government Information Quarterly</i> . 2013;30(1):S19–S31.
	Contracts	Sayogo DS, Pardo T. Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. <i>Government Information Quarterly</i> . 2013;30(1):S19–S31. Kim Y, Adler M. Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. <i>International Journal of Information Management</i> . 2015;35(4):408–18.
	Confidentially	Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B. The user's view on biodiversity data sharing—Investigating facts of acceptance and requirements to realize a sustainable use of research data <i>Ecological Informatics</i> . 2012;11:25–33. Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PLoS one</i> . 2015;10(2):e0118053. PMID:25714752 Kim Y, Adler M. Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. <i>International Journal of Information Management</i> . 2015;35(4):408–18. Schmidt B, Gemeinholzer B, Treloar A. Open data in global environmental research: The Belmont Forum's open data survey. <i>PLoS one</i> . 2016;11(1):e0146695. PMID:26771577

	Agreements	Sayogo DS, Pardo T. Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. <i>Government Information Quarterly</i> . 2013;30(1):S19–S31.
Data characteristics	Data size	Cragin MH, Palmer CL, Carlson JR, Witt M. Data sharing, small science and institutional repositories. <i>Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences</i> . 2010;368(1926):4023–38.
	Data format	Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PloS one</i> . 2015;10(2):e0118053. pmid:25714752 Cragin MH, Palmer CL, Carlson JR, Witt M. Data sharing, small science and institutional repositories. <i>Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences</i> . 2010;368(1926):4023–38. Ganzevoort W, van den Born RJ, Halfman W, Turnhout S. Sharing biodiversity data: citizen scientists' concerns and motivations. <i>Biodiversity and Conservation</i> . 2017:1–17.
	Data nature	Curry RG, Crowston K, Specht A, Grant BW, Dalton ED. Attitudes and norms affecting scientists' data reuse. <i>PLoS ONE</i> . 2017;12(12):e0189288. pmid:29281658
	Privacy	da Costa MP, Leite FCL. Factors influencing research data communication on Zika virus: a grounded theory. <i>Journal of Documentation</i> . 2019;75(5):910–26.
	Data quality	Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B. The user's view on biodiversity data sharing—Investigating facts of acceptance and requirements to realize a sustainable use of research data <i>Ecological Informatics</i> . 2012;11:25–33. Fecher B, Friesike S, Hebing M. What drives academic data sharing? <i>PloS one</i> . 2015;10(2):e0118053. pmid:25714752 Zuiderwijk A. <i>Open data infrastructures: The design of an infrastructure to enhance the coordination of open data use</i> . 's-Hertogenbosch: Uitgeverij BOXPress; 2015 Sayogo DS, Pardo T. Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. <i>Government Information Quarterly</i> . 2013;30(1):S19–S31. Yoon A, Kim Y. Social scientists' data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories. <i>Library & Information Science Research</i> . 2017;39(3):224–33.

Appendix C. Narrative Literature Review

Table 28 Narrative Literature Review

Authors	Title and year of publication	Study Objective	Publication	Keywords
Campbell	Access to scientific data in the 21st century: Rationale and illustrative usage rights review. (2015)	Understanding what drives users to use or not use the data from data repositories to ensure that the use of online open data is maximized.	Data Science Journal	'What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption'
Janssen, Charalabidis & Zuiderwijk	Benefits, Adoption Barriers and Myths of Open Data and Open Government. (2012)	Benefits and adoption barriers related to open data and open government are collected through research. In addition, some myths related to open data are explained to substantiate the reality.	Information Systems Management	Mail
Kitchin	The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. (2014)	The terms big data, open data and data infrastructures are explained. The data revolution is discussed, with possible consequences.	SAGE Publications Ltd.	Open data
Murray-Rust	Open Data in Science. (2008)	Explaining information about open data, why open data is important and how data should be made accessible to users.	Nat Prec	Open data
Piwowar & Vision	Data reuse and the open data citation advantage (2013)	Research into the patterns of data reuse based on the citation rate of researchers.	PeerJ	Mail
Zuiderwijk, Shinde & Jeng	What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. (2020)	A thematic analysis is conducted on the drivers and inhibitors of individual researchers. In addition, opportunities to test and use theories in open research data studies are outlined.	PLoS ONE	Mail
Weerakkody, Kapoor, Balta, Irani & Dwivedi	Factors influencing user acceptance of public sector big open data. (2017)	Examine the shortcomings in the field of big open data, especially the acceptance of big open data in the public sector.	Production Planning & Control	Open data sharing
Ryan & Deci	Intrinsic and extrinsic motivations: Classic definitions and new directions. (2000)	Both intrinsic and extrinsic motivation are outlined here. In addition, the relationships between these concepts are explained and further discussed what this does to people's daily lives.	Contemporary educational psychology	Intrinsic and extrinsic motivation
Tenkanen, Saarsalmi, Järvi, Salonen & Toivonen	Health research needs more comprehensive accessibility measures: integrating time and transportation modes from open data. (2016)	This study demonstrates the importance of transportation and temporality in the medical research field. The main focus is on accessibility and how this can be better understood through open data.	International journal of health geographics	Transportation domain open data
Mahajan, Kuehnel, Intzevidou, Cantelmo, Moeckel & Antoniou	Data to the people: a review of public and proprietary data for transportation models. (2022)	This study analyzes public data used within the transportation discipline and how useful this data is for that purpose. In addition, the data is classified according to how open this data is.	Transportation reviews	Transportation domain open data government
Zuiderwijk & Spiers	Sharing and re-using open data: A case study of motivations in astrophysics. (2019)	This research provides deep insight into the complex interaction between factors that influence the motivations of researchers to share or use open research data within the astrophysics discipline.	International Journal of Information Management	Mail
Zuiderwijk, Helbig, Gil-Garcia & Janssen	Special Issue on Innovation through Open Data - A Review of the State-of-the-Art and an Emerging Research Agenda: Guest Editors' Introduction. (2014)	The article is about the relationship between open data and innovations. It explores how open data activities lead to innovations.	Journal of theoretical and applied electronic commerce research	Zuiderwijk open data
Beer	Kwaliteit van combinatiebanen. (2022)	Research on multijobbers. The main focus is on the composition and quality of combination jobs.	Tijdschrift voor HRM	Extrinsieke en intrinsieke factoren
Baumeister & Leary	Writing Narrative Literature Reviews. (1997)	This article explains the narrative review, how to write it and its benefits	Review of General Psychology	Narrative literature review
Green, Johnson & Adams	Writing narrative literature reviews for peer-reviewed journals: secrets of the trade.(2006)	This report provides an overview of how the use of a narrative literature review	Journal of Chiropractic Medicine	Narrative literature review

Roopa & Rani	Questionnaire Designing for a Survey. (2012)	The definition of a questionnaire is explained. In addition, a method for checking a questionnaire for all necessary information is given.	Journal of Indian Orthodontic Society	Questionnaire method
-	Fields of Research (ANZSRC 2020): 3509 Transportation, Logistics and Supply Chains in Datasets - Dimensions. (z.d.)	This is one of the largest databases of scientific publications. Data can be searched by research discipline.	Dimensions	Research data sizes transportation
-	Publieke verkeers- en transportatiedata. (2020).	This report provides insight into how most traffic data is obtained within the Netherlands and how this data can in turn influence traffic situations.	Talking Traffic.	Transportation data
Schäfer, Pampel, Pfeiffenberger, Dallmeier-Tiessen, Tissari, Darby, Giaretta, Giaretta, Gitmans, Helin, Lambert, Mele, Reilly, Ruiz, Sandberg, Schallier, Schrimpf, Smit, Wilkinson & Wilson	Baseline Report on Drivers and Barriers in Data Sharing. (2011)	This study examines opinions about the Open Data Exchange. Stakeholders are asked to share their experiences and understanding of the topic.	Opportunities for Data Exchange (ODE)	Drivers open research data sharing
Corti, Eynden, Bishop & Woollard	Managing and Sharing Research Data: A Guide to Good Practice (Second). (2019)	This book provides information on how to handle primary and secondary data before, during and after a study. In addition, the book provides tips on using the right data and how to get the most out of the data in terms of effectiveness.	SAGE Publications Ltd.	Drivers open research data sharing
Feger, Pertiwi & Bonaiuti	Research Data Management Commitment Drivers: An Analysis of Practices, Training, Policies, Infrastructure, and Motivation in Global Agricultural Science. (2022)	This study investigates the drivers of researchers in the global agricultural sector on research data management.	Agricultural Research Knowledge	Drivers open research data sharing
European Commission	Transportation themes. Mobility and Transportation. (z.d.)	The subfields of the transportation sector are detailed here.	European Commission	Sub field of transportation
Huettmann	On the Relevance and Moral Impediment of Digital Data Management, Data Sharing, and Public Open Access and Open Source Code in (Tropical) Research: The Rio Convention Revisited Towards Mega Science and Best Professional Research Practices. (2015)	This article focuses on the basic principles of science and how these principles can lead to good decision-making.	Springer Science+Business Media	Data sharing relevance
Tedersoo, Küngas, Oras, Köster, Eenmaa, Leijen, Pedaste, Raju, Astapova, Lukner, Kogermann & Sepp	Data sharing practices and data availability upon request differ across scientific disciplines. (2021)	This article describes reasons why researchers do not openly share their research data.	Scientific Data	Data sharing disciplines
Tenopir, Allard, Douglass, Aydinoglu, Wu, Read, Manoff & Frame	Data Sharing by Scientists: Practices and Perceptions. (2011)	This article outlines the barriers to openly sharing research data. In addition, this article provides recommendations on how to share data more easily, as this is very important in today's science.	PLoS ONE	Data sharing disciplines
Morey, Forbath & Scoop	Customer data: Designing for transparency and trust. (2015)	This article focuses on the customers or people who give their data. This article outlines solutions to gain more trust by being transparent about how the data is sourced and used further.	Harvard Business Review	Transparency data
Taguchi	Description and explanation of pragmatic development: Quantitative, qualitative, and mixed methods research. (2018)	The methods quantitative and qualitative research methods are explained and a mix of them. This is explained in this article using examples of case studies.	System	Quantitative analysis explanation
Zenk-Möltgen, Akdeniz, Katsanidou, Naßhoven & Balaban	Factors influencing the data sharing behavior of researchers in sociology and political science. (2018)	This article explores how institutional and individual factors influence the open sharing of research data in sociology and political science.	Journal of Documentation	Influence data sharing
Yoon & Kim	Social scientists' data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories. (2017)	This research focuses on the reuse of data by researchers.	Library & Information Science Research	Data sharing disciplines
da Costa & Leite	Factors influencing research data communication on Zika virus: a grounded theory. (2019)	The study focuses on the factors that influence researchers to share research data concerns the Zika virus. Here, communication is the guiding principle on which the factors are based.	Journal of Documentation	Influencing factors data sharing

Enke, Thessen, Bach, Bendix, Seeger & Gemeinholzer.	The user's view on biodiversity data sharing— Investigating facts of acceptance and requirements to realize a sustainable use of research data Ecological Informatics. (2012)	This article explains what stopped the researchers from openly sharing the research data. It also provides solutions to help researchers share their research data more quickly.	Ecological Informatics	Data sharing motivations
Piowar, Day & Fridsma	Sharing detailed research data is associated with increased citation rate. (2007)	This study examines what the citation rate is for the research shared by the researcher.	PLoS ONE	Data sharing citation
Ganzevoort, van den Born, Halfman & Turnhout	Sharing biodiversity data: citizen scientists' concerns and motivations. Biodiversity and Conservation. (2017)	This article presents the motivations and resistances of biodiversity researchers to share research data	Biodiversity and Conservation	Data sharing motivations
Haeusermann, Greshake, Blasimme, Irdam, Richards & Vayena	Open sharing of genomic data: Who does it and why? (2017)	This article also focuses on open data sharing and researchers' motivations for sharing this research data.	PLoS ONE	Data sharing motivations
Joo, Kim & Kim	An exploratory study of health scientists' data reuse behaviors: Examining attitudinal, social, and resource factors. (2017)	This research focuses on exploring how health scientists are influenced by certain factors to reuse data.	Aslib Journal of Information Management	Data sharing
Kim & Adler	Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. (2015)	This article provides the individual, institutional and resource factors that social scientists influence in sharing research data.	International Journal of Information Management	Data sharing motivations
Sá & Grieco	Open data for science, policy, and the public good. (2016)	This article focuses on the benefits of open data sharing. This is supported by sharing a case study of the National Institute for Space from Brazil. Here the transition to an open data sharing environment is outlined.	Review of Policy Research	Open data science
Raffaghelli & Manca	Is there a social life in open data? The case of open data practices in educational technology research. (2019)	This article focuses on the transition from a closed data environment to open sharing of research data. This is outlined for several research fields.	Social Media and Open Science	Open data science
Wallis, Rolando & Borgman	If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. (2013)	The research is about the infrastructure within the data sharing platform and how this infrastructure can affect aspects related to research data sharing.	PLoS ONE	Data sharing infrastructure
Schmidt, Gemeinholzer & Treloar	Open data in global environmental research: The Belmont Forum's open data survey. (2016)	This research focuses on open data and all that comes with it. For example, it examines the expectations of a data sharing infrastructure and the barriers	PLoS ONE	Open data science
Zimmerman	Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. (2007)	This article examines what the requirements are for a good data sharing infrastructure in the ecological sector based on findings made before the data sharing infrastructure existed.	International Journal on Digital Libraries	Data reuse
Kim & Yoon	Scientists' data reuse behaviors: A multilevel analysis. (2017)	This study reviewed the factors that influence researchers in reusing research data. This was examined for several research disciplines.	Journal of the Association for Information Science and Technology	Data reuse behavior
Fecher, Friesike & Hebing	What drives academic data sharing? (2015)	This study outlines the process of data sharing from the researcher's point of view. It also provides solutions that lead to a greater likelihood of openly sharing research data.	PLoS ONE	Data sharing solutions
Ceci	Scientists' attitudes toward data sharing. (1988)	This study examines whether researchers are driven to share research data when they are financially rewarded for doing so. It also explores other solutions for encouraging open sharing of research data	SAGE Publications Ltd.	Data sharing solutions

Appendix D. The complete questionnaire

Opening statement

The ability and willingness to openly share research data differ per research discipline. The purpose of this questionnaire is to understand how extrinsic factors influence openly sharing research data in the transportation discipline. Completing this questionnaire will take you about 15 minutes.

The results of this study contribute to other projects and recommendations to facilitate open sharing of research data. In addition, the results of this research will be used for my master thesis research at TU Delft (master Complex Systems Engineering and Management) and may be published in a scientific journal. At the end of the questionnaire, you can indicate if you want to receive my thesis by e-mail.

Privacy protections and data storage

Your participation in this study is completely voluntary. You are also free not to answer questions. The data obtained will be stored securely within the European Union and will only be accessible by me, the researcher, and my two TU Delft supervisors. Your personal data will be removed six months after completing my study. By doing so, we minimize the risks in the event of a data breach.

By proceeding to the questionnaire, you agree with the above-mentioned statement.

Thank you in advance for completing this questionnaire.

In case of any questions, please contact the researcher;

Emily van den Hengel

Your Background (section 1/4)

This section asks for general information about your background.

Question 1: What is your nationality? (In the case of multiple nationalities, choose the nationality with which you feel most connected)

- Drop down list (Qualtrics)

Question 2: Which country are you currently living in? (if you live in multiple countries, choose the country where you spend most of your time)

- Drop down list (Qualtrics)

Question 3: Which of the following best describes your current role?

- I am a full professor
- I am an associate professor
- I am an assistant professor
- I am a postdoctoral researcher
- I am a PhD candidate
- I am a researcher at a university (not being a professor, postdoc, or PhD candidate)
- I am a researcher at a research institute
- I am a researcher at a company
- Other, namely: ...

Question 4: In what research field are you primarily active? (Multiple choice)

- Mobility Strategy
- Passenger rights
- Security and safety
- Clean transportation
- Urban transportation
- Sustainable transportation
- Infrastructure and investment
- Intelligent transportation systems
- Research and innovation
- International relations
- Public service obligations
- Transportation of dangerous goods
- Logistics and multimodal transportation
- Other, namely: ...

Question 5: Would you say that your research field has specific characteristic that influence the ease or difficulty of openly sharing research data?

- Yes
 - o If yes, which ones?
- No

Your research data (section 2/4)

This section focuses on the research data you are working with.

Question 6: What actions do you perform with research data as a part of your daily occupation? Research data concerns any data related to research, including observational, theoretical, empirical, and other data. It concerns both primary and secondary data.

- Creating research data
- Processing data
- Analysing data
- Preserving data
- Giving access to data
- Re-using data
-

Question 7: What is the size of the research data you typically work with? (multiple choice)

- < 100 GB
- 100 GB – 1 TB
- > 1 TB
- I don't know

Question 8: What data format is most common for the data you work with?

- PDF
- Image scans
- Excel
- CSV
- RDF
- SPARQL
- All the above
- Other
- I don't know

Question 9: What is the nature of the research data you primarily work with?

- Qualitative nature
- Quantitative nature
- Both

Question 10: How often is your research funded by a third party? (Third parties are institutions that influence research from the outside. Examples are; universities, governments, public institutions and private institutions)

- 0% of the time
- 0% - 20% of the time
- 20% - 40% of the time
- 40% - 60% of the time
- 60% - 80% of the time
- 80% - 100% of the time

→ Go to next section

Your experience with open research data (section 3/4)

Research data that is openly available to anyone on the internet is called "open research data". This data should be freely accessible, adaptable, reusable, and shareable with other researchers. Open research data refers to data obtained from both quantitative and qualitative studies.

Question 11: In the past 5 years, have you ever openly shared your research data? In this study, we define openly sharing research data as publishing research data on a digital platform, portal, or repository, which is openly accessible to anyone for free.

- Yes (if yes, go to question 11)
- No (if no, go to question 16)

Question 12: What type of data did you share openly in the past 5 years?

- Quantitative data
- Qualitative data
- Both quantitative and qualitative data
- I don't know

Question 13: Which share of your data collections have you openly shared your research data in the past 5 years?

- 0% - 20% of the data
- 20% - 40% of the data
- 40% - 60% of the data
- 60% - 80% of the data
- 80% - 100% of the data

Question 14: Through what types of repositories did you openly share your research data in the past 5 years?

- An institutional/ university data repository
- A domain-specific repository
- A journal's data repository
- Other, namely ...

Question 15: What were your reasons for openly sharing research data in the past 5 years?

- Requirements from funding agency/agencies
- Requirements for publishing a journal article
- More recognition from supervisors, colleagues, peers or others
- More citations by peers and others
- For learning and training purposes
- To set up follow-up research and/or collaborations
- Other, namely...

→ Go to next section

Question 16: What have been the biggest barriers in not sharing research data over the past 5 years?

- The data are too (privacy) sensitive
- One of the organizations I work with does not want me to share my data openly.
- I don't know where to share my open research data.
- I have no confidence in current data repositories.
- I fear that my data would be criticized.
- My organization does not require me to openly share my data.
- Openly sharing data is very time-consuming.
- I have no experience sharing open research data
- Other, namely...

→ Go to next section

Statements on extrinsic factors affecting research data sharing (section 4/4)

In the following sections, we ask you to indicate to what extent the presented extrinsic factors influence your decision to openly share your research data or not. Note that the statement concern your non-privacy sensitive data.

→ Go to next section

Requirements and formal obligations

Requirements and formal obligations are agreements made with third parties. These agreements must be fulfilled by the researcher. Often these agreements have to do with getting funding from a third party (e.g., national science foundations, the European Commission, or industry) in exchange for a specific action with the research data or research. This may include sharing the research data or not.

Question 16: To what extent do the following requirement-related factors encourage you to openly share your (non-privacy sensitive) research data or not?

Requirements and formal obligations	Discourage me from openly sharing my research data	Slightly discourage me from openly sharing my research data	Do not discourage me from or encourage me to openly share my research data	Slightly encourage me to openly share my research data	Encourage me to openly share my research data
Funding of my research by funding agencies					
Mandates and compliances policies by third parties I collaborate with					
The encouragement by journals to openly share my research data					
Clear codes of ethics within my research field					

→ Go to next section

Legislation and regulation

Legislation and regulation are aspects to which the researcher must adhere. It involves public policy and its laws and regulations. So it is not about agreements made with third parties, but about agreements that everyone, who falls within the field of legislation, must adhere to.

Question 17: To what extent do the following legislation and regulation-related factors encourage you to openly share your (non-privacy sensitive) research data or not?

Legislation and regulation	Discourage me from openly sharing my research data	Slightly discourage me from openly sharing my research data	Do not discourage me from or encourage me to openly share my research data	Slightly encourage me to openly share my research data	Encourage me to openly share my research data
Data ownership protection / Ownership laws and regulations					
Scientific integrity protection / Scientific integrity laws and regulations					
Data protection / privacy laws and regulations (e.g., the General Data Protection Regulation)					
The requirements to create a Research Data Management Plan					

→ Go to next section

Facilitating conditions

Various conditions may facilitate openly sharing research data.

Question 18: To what extent do the following facilitating conditions-related factors encourage you to openly share your (non-privacy sensitive) research data or not?

Facilitating conditions	Discourage me from openly sharing my research data	Slightly discourage me from openly sharing my research data	Do not discourage me from or encourage me to openly share my research data	Slightly encourage me to openly share my research data	Encourage me to openly share my research data
Accessible research data repositories, such as 4TU.ResearchData and DANS					

A supportive (technical) infrastructure of research data repositories					
Support provided by research data repositories (e.g., checking privacy compliance and metadata support)					

➔ Go to next section

Trust

Various trust-related factors may influence openly sharing research data.

Question 19: To what extent do the following trust-related factors encourage you to openly share your (non-privacy sensitive) research data or not?

Trust	Discourage me from openly sharing my research data	Slightly discourage me from openly sharing my research data	Do not discourage me from or encourage me to openly share my research data	Slightly encourage me to openly share my research data	Encourage me to openly share my research data
A data repository that provides information about various quality indicators of the data (e.g., completeness, timeliness, consistency)					
A data repository that provides sufficient metadata to allow the interpretation of the data					
A data repository that allows data providers and data users to interact with each other					
A data repository that guarantees the storage and availability of the data for at least the next 10 years					
A data repository provided by one or more organizations I know					
A data repository using data standards that are accepted in my field					

➔ Go to the next section

Expected performance

Expected performance has to do with both the research itself and the reactions of the outside world.

Question 20: To what extent do the following expected performance-related factors influence your decision to openly share your (non-privacy sensitive) research data or not?

Expected performance	Discourage me from openly sharing my research data	Slightly discourage me from openly sharing my research data	Do not discourage me from or encourage me to openly share my research data	Slightly encourage me to openly share my research data	Encourage me to openly share my research data
The possibility that my research will more visible to a wider audience					
The possibility that my research would be criticized					
The possibility that my research would be cited more often					
The data is potentially useful for other researchers					
Collaboration with other researchers					
A reward I could receive for openly sharing research data					

A punishment I could receive for not openly sharing my research data					
--	--	--	--	--	--

→ Go to the next section

Social influence and affiliation

Various social influence-related factors may influence openly sharing research data.

Question 21: To what extent would the following social influence and affiliation-related factors encourage you to openly share your (non-privacy sensitive) research data or not?

Social influence and affiliation	Discourage me from openly sharing my research data	Slightly discourage me from openly sharing my research data	Do not discourage me from or encourage me to openly share my research data	Slightly encourage me to openly share my research data	Encourage me to openly share my research data
If my supervisor would tell me I should openly share my research data					
If my colleagues would tell me I should openly share my research data					
If peers from my field would tell me I should openly share my research data					

→ Go to the next section

Effort

Various effort-related factors may influence openly sharing research data.

Question 22: To what extent do the following effort-related factors encourage you to openly share your (non-privacy sensitive) research data or not?

Effort	Discourage me from openly sharing my research data	Slightly discourage me from openly sharing my research data	Do not discourage me from or encourage me to openly share my research data	Slightly encourage me to openly share my research data	Encourage me to openly share my research data
Easy to use repositories for openly sharing research data					
Considerable time that needs to be invested to openly share research data					
The considerable effort it costs to openly share research data					
A data steward whom I can ask questions about openly sharing research data					

→ Go to the next section

Researcher's experience and skills

This category focuses on whether the purpose of the research data, such as training or learning, affects the sharing of the research data.

Question 23: To what extent do the following experience and skills-related factors influence your decision to openly share your (non-privacy sensitive) research data or not?

Researcher's experience and skills	Discourage me from openly sharing my research data	Slightly discourage me from openly sharing my research data	Do not discourage me from or encourage me to openly share my research data	Slightly encourage me to openly share my research data	Encourage me to openly share my research data

Experience with openly sharing research data					
Me having the skills to openly share research data					
Me having received training or education on how to openly share research data					

→ Go to the next section

Data characteristics

Various data characteristics-related factors may influence openly sharing research data.

Question 24: To what extent do the following data characteristics-related factors encourage you to openly share your (non-privacy sensitive) research data or not?

Data characteristics	Discourage me from openly sharing my research data	Slightly discourage me from openly sharing my research data	Do not discourage me from or encourage me to openly share my research data	Slightly encourage me to openly share my research data	Encourage me to openly share my research data
My research data being a large data size					
My research data being a small data size					
My research data being stored in a commonly used format in my field					
My research data being qualitative					
My research data being quantitative					
My research data being high quality					
My research data being low quality					

→ Go to next section

Closure

Thank you for completing my questionnaire. I appreciate it tremendously.

Question 28: Do you have any other questions or comments? If so, please leave them here.

→ Go to next section

I thank you for your time spent taking this survey.

Your response has been recorded

If you are interested in my thesis, please click here to leave your e-mail address. I will make sure to send you my master's thesis when it is completed.

Appendix E. HREC approval

Date 13-Dec-2022
Contact person Dr. Cath Cotton, Policy Advisor Academic Integrity



Human Research Ethics
Committee TU Delft
(<http://hrec.tudelft.nl/>)
Delft
The Netherlands

*Ethics Approval Application: Open data sharing in the transport sector
Applicant: Hengel, Emily van den*

Dear Emily van den Hengel,

It is a pleasure to inform you that your application mentioned above has been approved.

Good luck with your research!

Sincerely,

Dr. Ir. U. Pesch
Chair HREC
Faculty of Technology, Policy and Management

Figure 7 HREC approval

Appendix F. Additional results

Background information

Table 28 lists all the characteristics of the respondents. The nationality, country of residence, occupation, current position, and research fields of all respondents can be found in this table. These figures are explained further in this section.

Table 29 Background figures

Demographic	Aspects	Percentage (frequency)
Nationality (N=77)	European	72.7% (56)
	South American	7.8% (6)
	Asian	18.2% (14)
	African	1.3% (1)
Country living in (N=77)	Europe	85.7% (66)
	South America	2.6% (2)
	Asia	11.7% (11)
Occupation (N=77)	Academic	92.2% (71)
	Non-academic	7.8% (6)
Current positions (N=77)	Full professor	14.3% (11)
	Associate professor	13.0% (10)
	Assistant professor	10.4% (8)
	Postdoctoral researcher	5.2% (4)
	PhD candidate	33.8% (26)
	Researcher at university	5.2% (4)
	Researcher at research institute	1.3% (1)
	Researcher at company	5.2% (4)
Research fields (N=69)	Mobility Strategy	30.4% (21)
	Security and Safety	5.8% (4)
	Clean Transportation	11.6% (8)
	Urban Transportation	43.4% (30)
	Sustainable Transportation	46.4% (32)
	Infrastructure and Investment	11.6% (8)
	Intelligent Transportation Systems	20.3% (14)
	Research and Innovation	15.9% (11)
	International Relations	2.9% (2)
	Public Service Obligations	7.2% (5)
Usage of data (N=60)	Transportation of Dangerous Goods	1.4% (1)
	Logistics and Multimodal Transportation	15.9% (11)
	Other	17.4% (12)
	Creating research data	58.3% (35)

	Processing data	75.0% (45)
	Analyzing data	95.0% (57)
	Preserving data	26.7% (16)
	Giving access to data	25.0% (15)
	Re-using data	61.7% (37)

Characteristics of the research fields

The table below shows whether respondents felt that the transportation discipline contains certain influencing characteristics that influence the open sharing of research data. Just under half of the respondents indicated that there are influencing characteristics within the transportation discipline that influence the open sharing of research data.

Table 30 The influencing characteristics of the transportation discipline on open data sharing

Influencing characteristics on open data sharing			
Transportation discipline	Yes	No	Percentage 'Yes' within the research field
	32	35	47.8%

Data sharing characteristics

To distinguish between respondents who openly share research data and researchers who do not, this question was added to the questionnaire. Based on the respondents' answer, follow-up questions were asked to contribute to the study. The rationale for sharing or not sharing research data was revealed. Some of these motivations are part of the extrinsic factors tested to answer the research question.

56 respondents answered whether they had shared research data in the past 5 years. More than half have not shared research data in the past 5 years. 41.1% of respondents, or 23 respondents, have shared research data on occasion in the past 5 years. The figure below outlines the exact numbers.

Table 31 Data sharing characteristics

Demographic	Aspects	Percentage (frequency)
Sharing data in the past 5 years (N=56)	Yes	41.1% (23)
	No	58.9% (33)
Data nature shared the past 5 years (N=23)	Qualitative nature	13.0% (3)
	Quantitative nature	47.8% (11)
	Both	34.8% (8)
	I don't know	4.3% (1)
Share of data shared (N=23)	0% - 20% of the data	21.7% (5)
	20% - 40% of the data	17.4% (4)
	40% - 60% of the data	17.4% (4)
	60% - 80% of the data	21.7% (5)
	80% - 100% of the data	21.7% (5)

Table 32 Data sharing characteristics

Nationality and country of residence of the respondents

The table below outlines the nationalities relative to the countries in which they live.

Table 33 Nationalities and countries of residence of the respondents

Nationality	Country living in																			
	Austria	Bahrain	Belgium	Brazil	China	France	Germany	Greece	India	Ireland	Israel	Netherlands	Norway	Slovenia	Spain	Sweden	Switzerland	Turkey	United Kingdom	
Afghanistan	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Algeria	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Argentina	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Austria	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Belgium	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
Brazil	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Chile	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
China	0	0	0	0	2	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
Colombia	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
France	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Germany	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1
Greece	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
India	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0
Iran	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
Ireland	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Israel	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
Italy	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2
Latvia	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Netherlands	0	0	0	0	0	0	0	0	0	0	1	23	0	0	0	0	1	0	1	1
Norway	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Slovenia	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Spain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Sweden	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Turkey	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
United Kingdom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9

Current positions and the locations of the respondents

The table below shows the distribution regarding the respondents' current positions and where they are currently in the world.

Table 34 The academic roles of the respondents per country of living

Country living in	Current role									
	Full professor	Associate professor	Assistant professor	Postdoctoral researcher	PhD candidate	Researcher at university	Researcher at research institute	Researcher at company	Other	
Austria	0	0	1	0	1	0	0	1	0	
Bahrain	1	0	1	0	0	0	0	0	0	
Belgium	0	0	0	0	1	0	0	0	0	
Brazil	1	1	0	0	0	0	0	0	0	
China	1	0	0	0	1	0	0	0	0	
France	0	0	0	0	2	0	0	0	0	
Germany	0	0	0	0	1	0	0	1	0	
Greece	0	0	0	0	1	0	0	0	0	
India	0	0	1	0	1	0	0	0	0	
Ireland	0	0	0	0	0	0	0	0	1	
Israel	1	0	0	1	0	0	0	1	0	
Netherlands	6	4	5	1	17	1	0	0	3	
Norway	0	0	0	0	1	0	0	0	0	
Slovenia	1	0	0	0	0	0	0	0	0	
Spain	0	0	0	1	0	0	0	0	0	
Sweden	0	0	0	0	0	0	0	1	0	
Switzerland	0	0	0	0	0	0	0	0	1	

Turkey	0	1	0	0	0	0	0	0	0	0
United Kingdom	0	4	0	1	0	3	1	0	4	

Influencing characteristics per research field on sharing research data

The table below shows the comprehensive distribution regarding the different research fields and whether these respondents have shared research data in the past 5 years.

Table 35 The influencing characteristics of the research field on open data sharing

Influencing characteristics on sharing research data				
Field of Research		Yes	No	Percentage 'Yes' within the research field
	Mobility Strategy	10	11	47.6%
	Security and Safety	3	1	75.0%
	Clean Transportation	4	4	50.0%
	Urban Transportation	13	17	43.3%
	Sustainable Transportation	13	18	41.9%
	Infrastructure and Investment	6	2	75.0%
	Intelligent Transportation Systems	9	5	64.3%
	Research and Innovation	7	4	63.6%
	International Relations	0	2	0.0%
	Public Service Obligations	1	4	25.0%
	Transportation of Dangerous Goods	0	1	0.0%
	Logistics and Multimodal Transportation	8	3	72.3%
	Other, ...	5	6	45.5%
	<ul style="list-style-type: none"> • Air Traffic Management • Climate Risk • Freight Transportation and Trade • Public Space • Reverse Logistics • Transportation Network Modelling • Transportation Planning • Travel Behavior • Urban Planning 			

Table 36 The influencing characteristics of the research field on open data sharing

Data characteristics

Data size allows the respondent to indicate what data size they normally work with.

Table 37 The data sizes respondents work with

Data size	Frequency of answers chosen		Percent among respondents	
< 100 GB	41	68.3%		
100 GB – 1 TB	10	16.7%		
> 1 TB	3	5.0%		
I don't know	10	16.7%		

A chi-square test was performed for data size. This shows that for three of the four response options, it is found that they do not meet the conditions of this test. Therefore, it was chosen to analyze all analyses in the

same way, namely by the Fisher's Exact Test. First, the table below outlines how respondents answered regarding data size and sharing.

Table 38 Distribution of data size relative to data sharing

Shared Data				
Data size N = 56		Shared data the past 5 years [Yes No]		P-value
	<100GB	19	20	0.139
	100GB – 1TB	3	6	0.723
	>1TB	1	2	1.000
	I don't know	1	8	0.067

The results from the Fisher's Exact Test can be found in table 37 In particular, the response option 'I don't know' and the response option '<100GB' have a good probability of having a link with the variable 'sharing or not sharing the research data'.

First, the formats used by respondents were analyzed. For this analysis, 59 respondents were analyzed. Again, it was possible for respondents to indicate multiple data formats. The table below outlines all the figures, along with the percentage that chose that data format divided by the total number of respondents.

Table 39 The data formats respondents work with

	Frequency of answers chosen		Percent among respondents
Data formats	PDF	18	30.5%
	Image scans	4	6.7%
	Excel	35	59.3%
	CSV	45	76.3%
	RDF	2	3.4%
	SPARQL	2	3.4%
	All of the above	1	1.7%
	I don't know	2	3.4%
	Other, ...	18	30.5%
	<ul style="list-style-type: none"> • Text files • Audio • Database file • Geopackage • GeoTIFF • Shapefile • Parquet • Gis files • GPKG files • JSON files • MS Word files • SPSS files • STATA files • Python files • XML files • MTS files 		

The data format distinguishes between different formats used by respondents to process the data in. Table 39 lists the exact numbers of respondents who use a particular data format against whether or not they share data.

Table 40 Distribution of data format relative to data sharing

Shared Data				
Research field N = 56		Shared data the past 5 years [Yes No]		P-value
	PDF	9	7	0.144
	Image scans	2	1	0.365
	Excel	14	18	0.638
	CSV	22	22	0.009
	RDF	0	2	0.343
	SPARLQ	2	0	0.164
	All the above	0	1	0.589
	Other	8	9	0.548
	I don't know	0	2	0.343

Three of the nine response options can be analyzed with the chi-square test. The remaining six were analyzed with the Fisher's Exact Test. First, the calculations with the chi-square test are outlined. Enough respondents named "PDF" in response to the question of what data format they use. This ensures that the sample is large enough for a chi-square test. All conditions are met. The table below outlines the figures regarding the chi-square test. Again, the p-value is considered continuous. The p-value is 0.144. This means that it is valid that the data format 'PDF' has a connection with whether or not the data is shared.

Another data format on which a chi-square test can be tested is Excel. All conditions were met and thus the chi-square test was used to obtain results. The p-value of the data format "Excel" tested with whether or not the research data were shared is reasonably high, 0.638. The probability that there is a connection between these two variables is not significant. However, it can never be completely ruled out.

The final data format for which a chi-square test is used is the response option "Other. Respondents had the option to choose this answer if they used other formats not covered by the previous answers. Quite a few respondents chose this answer option. All frequencies are higher than the value of 5, so all conditions were met. The p-value between the data format 'Other' and the dependent variable related to sharing or not sharing the research data is 0.548. This p-value is significantly high making it unlikely that there is a relationship between these two variables.

The other response options are analyzed using the Fisher's Exact Test. This test is performed when one of the conditions of the chi-square test is not met due to, for example, too small a number of respondents. Table 38 sets out all the figures related to the Fisher's Exact Test.

The response option 'CSV' with a p-value of 0.010 has a significant possibility of relationship between the data format 'CSV' and the dependent variable that focuses on sharing or not sharing the research data. Even according to the Null Hypothesis Significance Testing theory, the null hypothesis is rejected, and it can be said that there is a connection (McShane et al., 2019). However, the p-value is considered a continuous variable, where; the smaller the number the stronger the evidence emerging from the testing.

The response option 'SPARLQ' with a p-value of 0.164 has a significant possibility of a connection between the data format 'SPARLQ' and the variable that focuses on sharing or not sharing the research data.

The remaining data formats have a higher p-value that significantly reduces the likelihood of a connection between the variables.

The nature of the data can be distinguished into quantitative and qualitative. Respondents were asked to indicate what kind of data they often work with. They may also work a lot with both types of data. This can also be indicated.

Table 41 The nature of the data respondents work with

		Frequency of answers chosen	Percent among respondents
Data nature	Qualitative nature	5	8.5%
	Quantitative nature	33	55.9%
	Both	21	35.6%

Table 41 outlines the numbers of respondents who use a particular data nature against whether or not they share the research data.

Table 42 Distribution of data nature relative to data sharing

Shared Data				
Data nature N = 56		Shared data the past 5 years [Yes No]		P-value
	Qualitative nature	2	3	1.000
	Quantitative nature	14	16	0.361
	Both	7	14	0.362

Two of the three response options can be analyzed with the chi-square test. These are the answers "quantitative" and "both. The answer 'qualitative' was analyzed with the Fisher's Exact Test because it does not meet the conditions of the chi-square test. The p-value for the quantitative data is 0.361. This value could indicate that there is a connection between the quantitative data and the sharing or not sharing of the research data.

This p-value of the response option "both" is almost the same as for the quantitative data. The p-value is 0.362. So again, there may be a chance that there is a relationship between the dependent and this variable.

The qualitative data is tested with the Fisher's Exact Test. The number of respondents who chose this answer is not large enough for a chi-square test and therefore all conditions are not met. With the value for the 2-sided significance, or p-value, of 1.000, it is very unlikely that there is a relationship between the use of qualitative data and whether or not the survey data is shared. One reason for this may be that there were too few respondents to test this properly.

Researchers may face funded investigations by third parties. Consider parties such as, universities, governments, public and private institutions. These parties often make it possible for research to be conducted. In exchange for the funding, third parties may have established certain requirements that the researcher must adhere to. These requirements may affect the open sharing of research data. Therefore, it is important to know the extent to which respondents have to deal with third-party funding. That way it is clear how much influence, if any, they have on the open sharing of research data. For this analysis, 58 respondents were included. Respondents answered how often the research was funded by a third party.

Table 43 Research funded by a third party

		Frequency of answers chosen	Percent among respondents
Funded research	0% of the time	7	12.1%
	0% - 20% of the time	9	15.5%
	20% - 40% of the time	4	6.9%

40% - 60% of the time	8	13.8%
60% - 80% of the time	7	12.1%
80% - 100% of the time	23	39.7%

Respondents can indicate how often they are funded by a third party to conduct a research. This may influence whether or not they share research data.

Table 44 Distribution of research funded relative to data sharing

Shared Data				
Research funded by a third party N = 56		Shared data the past 5 years [Yes No]		P-value
	0% of the time	2	5	0.688
	0% - 20% of the time	3	6	0.723
	20% - 40% of the time	1	2	1.000
	40% - 60% of the time	2	6	0.449
	60% - 80% of the time	4	2	0.215
	80% - 100% of the time	11	12	0.421

Five of the six response options have too small a number of respondents to meet the conditions of the chi-square test. As a result, a Fisher's Exact Test was performed for all response options.

The figure above shows that all p-values are on the high side. However, the p-values associated with the responses '40%-60% of the time', '60%-80% of the time', and '80%-100% of the time' are the lowest. For these values, there is still a chance that there is a connection between these independent variables and the dependent variable that focuses on sharing or not sharing the research data.