



Regulation of open access to research data

**A study about open access to research data and
the role of the Dutch government**

© S.L. Meijer

Brussels – Belgium | 2014

Master thesis Comparative Public Administration

Radboud University Nijmegen





Regulation of open access to research data

A study about open access to research data and the role of the Dutch government

Brussels – Belgium | 2014

Lecturer Radboud University Nijmegen:

Prof. dr. S. van Thiel

Second lecturer Radboud University Nijmegen:

dr. J. H.M. M. Tholen

Lecturer Neth-ER:

drs. S. W. den Bak

Student:

S.L. Meijer

Student number:

4061241

Faculty:

Faculty of Management Sciences

Study:

Comparative Public Administration

Date:

October 1, 2014

Signature:

Radboud University Nijmegen



Preface

Open access does give opportunities, just like finishing this master thesis. I am grateful that I was able to study at the university. By finishing this thesis, I will receive my masters degree on Comparative Public Administration at the Radboud University Nijmegen. Finalising my study period will open up new opportunities, but will also end my great student life. I am really grateful for the opportunities I had during my studies and the personal development during that period.

The topic of this thesis was unknown to me before starting this study. By doing this study I improved my perspective on information management and governmental regulation. It is a upcoming debate in science and society, what made this study more interesting to work on. The thesis is mainly written at the office of Neth-ER, the representation of the Dutch knowledge field in Brussels. This period widened my European perspective and the relations with the Netherlands, also because I was living in Brussels for six months. Knowledge and research is important for society because research can improve (the processes in) society. The topic of this research is an example of that aim to improve and boost research and society.

The realisation of this thesis was not always easy, mainly because of the deviating structure of this thesis. A lot of people supported me during this research and I would like to say thank you those people. First of all, I would like to thank my lecturer from the Netherlands, Sandra van Thiel. She gave me constructive and fast feedback during my stay in Brussels and the meetings at the university. Her approach is very pleasant, because she is friendly, structured and to the point. Secondly, my lecturer Sebastiaan den Bak. He supervised me during my internship in Brussels and he gave helpful personal and substantive advice. Thirdly, I want to thank all my colleagues (Fried, Edith, Ingrid, Karlijn, Kimberly, Charlotte, Tim and Marleen) at Neth-ER for supporting my thesis process, and also for the great and informal working atmosphere. Fourthly, I want to thank my fellow students (Bart, Eduard, Lotte, Perihan and Steven) that participated in the feedback sessions at the university by giving advice on my thesis. I also want to thank my experts in the experts panel (Mrs. Bruce, Mr. De Nooijer, Mrs. Dillo, Mr. Madho, Mrs. Meyer, Mr. Rombouts, Mr. Sondervan, Mrs. Van Nieuwerburgh and Mr. Wittenburg) for their enormous input and their time for answering all the questions. I want to thank my interviewees for their advice and time during this research. In particular, Mrs. Ridder-Numan, Mr. Van Loenen, Mr. Van den Biesen and Mr. Spichtinger. Also many thanks to the other interviewees for their information: Mr. Hof, Mr. Grosfeld, and Mr. Kolman. Last but not least, I want to thank my family and especially my girlfriend Leonieke for supporting me during this thesis process. The time with my family and friends in Brussels was a great advantage of doing this thesis abroad. This thesis widened my perspective which has been a great opportunity.

Summary

Accessibility to research data is an upcoming issue in science and society. The possibility to have access to research data is part of the 'open' movement of current society. Open access to research data is the right to access and re-use digital research data, without any additional costs for the user. The general opinion is that research data need to be accessible for the public. People want to have access to data, because people feel they have the right to access the data; it is about 'democratisation of data'. 'Open by default' is the central aim of open access to research, by taking into account sensitive and commercial data and the interests of businesses and industry.

Open access to research data creates opportunities. Easier access to research data can contribute to poor efficiency of research data by re-using data and increases the quality of research. Open access facilitates validation, verification and evaluation of research (results). However, open access to research data does have down-sides. Issues related to privacy, confidentiality or security could form a barrier for open access. Furthermore, there is a lack of awareness for the benefit that open research data could bring to the science system. Also the infrastructure of data systems and other technical aspects of data depositing, storage and security are not always well-known for researchers. Costs of publishing and maintaining data can also hinder open access to research data.

In general, there is agreement on the necessity of intervention and a regulatory framework on open access to research data. The focus of the framework should be: 'as open as possible, closed if needed'. A regulatory framework should at least be flexible, discipline dependent and approached on a case-by-case basis. A data management plan is recommended and should at least entail information about: discoverability, protection of data, data storage and data authenticity. The framework should create an environment of awareness, trust and recognition and support for researchers. The feasibility of such a regulatory framework depends on the willingness and commitment of all stakeholders. Therefore a continuous open dialogue between the government and all stakeholders should be set up. The government should take up a coordinating role by initiating meetings and training for all stakeholders. Additionally, the government should stay informed and should inform stakeholders. Facilitation is also proposed by giving financial support to researchers. The government should also take up a role as legislator by creating soft law. Incentives could also be implemented, like encouraging institutional recognition of publishing data for researchers. In general, governmental intervention should only be done if necessary. Regulation can also be done at other levels (by institutions, publishers and funders). In addition, regulation should be done at several different levels and in parallel. Hybrid regulation and cooperation in open access to research data is necessary in order to boost research and innovation.

Table of contents

Chapter	Page
Preface	3
Summary	4
Table of contents	5
1. Introduction	
1.1 General context	8
1.2 Problem definition	9
1.3 Relevance for society and science	9
1.4 Outline of the report	10
2. Theory part one	
2.1 General context	12
2.2 Legislation	15
2.3 European perspective	15
2.3.1 Open access to scientific publications	16
2.3.2 Open access to research data	16
2.4 Policy instruments	18
2.4.1 Typologies	18
2.4.2 Hybrid regulation	23
2.4.3 Digital era	24
2.5 Summary	25
3. Research methodology	
3.1 Notification	27
3.2 Sensitizing concepts	27
3.3 Research method	28
3.3.1 Delphi in-depth	30
3.4 Research quality	32
3.5 Limitations	33
3.6 Summary	35

Chapter	Page
4. Analysis	
4.1 Notification	36
4.2 Definition	37
4.2.1 Research data	37
4.2.2 Open access	40
4.2.3 Sensitive data	43
4.2.4 Commercial data	44
4.3 A solution to what?	46
4.4. Necessity	49
4.4.1 What if not?	51
4.5 Ownership	53
4.6 Opportunities	56
4.7 Difficulties	58
4.8 Intervention	61
4.8.1 Governmental level	63
4.8.2 Institutional level	65
4.8.3 Funders level	66
4.8.4 Publishers	67
4.9 Framework	67
4.9.1 Criteria	68
4.9.2 Data management	73
4.9.3 Data authenticity	75
4.9.4 Data storage	76
4.9.5 Finance and maintenance	79
4.9.6 Licensing scheme	81
4.9.7 Protecting data	82
4.9.8 Framework feasibility	83
4.10 Incentives for researchers	84
4.11 Business and industry	86
4.11.1 Philips perspective	88
4.12 International perspective	91

Chapter	Page
5. Theory part two	
5.1 Arguments for intervention	97
5.2 Detectors and effectors	98
5.3 Sticks, carrots and sermons	100
5.3.1 Other regulatory levels	102
5.4 Strategies	104
5.5 Hybrid regulation	105
5.6 Digital era	107
5.7 Summary	108
6. Conclusions and recommendations	
6.1 Research questions	110
6.2 Reflection	119
6.3 Recommendations future research	120
References	122
Appendix A List of Delphi-analysis experts	131
Appendix B List of interviewees	132
Appendix C Open research data pilot areas	133
Appendix D BusinessEurope position paper	134
Appendix E Philips position paper	137
Appendix F Guidelines Data Seal of Approval	139
Appendix G Legal and ethical issues	140
Appendix H Interview topic list	143
Appendix I Delphi-analysis questions	145

1. Introduction

This chapter presents an overview of the study. It contains information about the aim and structure of the study (1.1). The problem definition will be presented (1.2), as well the relevance of this study (1.3). This chapter concludes with an outline of the whole report (1.4).

1.1 *General context*

Openness of information is a main topic in discussions in science and politics. It has become a common concept in a growing number of scientific and academic fields (Guibault & Wiebe, 2013). Open access to information lies at the core of the discussion. Open access entails the accessibility to scientific publications or research data. This study discusses the main issues about the open access to research data. Research data are the evidence for the published amount of scientific knowledge (Buchholtz, Bukowski & Śniegocki, 2013; OECD, 2007). Opening this collection of information is the new debate after open access to scientific publications, which has already been implemented in many countries (Björk, Welling, Laakso et al., 2010; Guibault & Wiebe, 2013; Schmidt & Kuchma, 2012).

Open access issues have a deep effect on digital information (García-Peñalvo, García de Figuerola & Merlo, 2010). According to the European Commission, scientific publications are no longer the only elements on the way to open access: research results upon which publications are based must also be made available to the public (Guibault, 2013).

The access to data can be restricted if an individual or institution while looking for evidence and resources like papers and reports has to use codes and or pay for repositories or databases. The accessibility to research data is also lacking, when information is not available to the general public.

Accessibility to research data could be very important for the acceleration of research dissemination and uptake of research findings (Correia & Teixeira, 2005, p. 353; Eysenbach, 2006; García-Peñalvo, García de Figuerola & Merlo, 2010; Guibault, 2013). Moreover, it enhances the production and validation of knowledge (Guibault & Wiebe, 2013). However, the impact of openness is part of a large social debate, mainly because its legal implications are unknown (Guibault & Wiebe, 2013). The growing demand for open access to information causes new opportunities, but also creates challenges in the digital world. Issues such as privacy of individuals and the growing amount of data asks for coordination or regulation.

So this study starts the debate of open access to research data by discussing the current situation of open access to research data. The role of the government will be discussed as regulator in the broadest sense. Experts will discuss the problems and possible steps in the process of open

access to research data. This will lead to recommendations on an appropriate way of regulation the issue of open data.

1.2 *Problem definition*

The aforementioned developments lead to the following problem definition, which consists of a research goal and research questions.

The aim is based on the aforementioned developments and therefore the aim of this study can be described as:

Prescribe if and how the Dutch government can regulate open access to research data.

The general question is:

To what extent are possibilities available and desirable for the Dutch government to regulate open access to research data?

Sub questions organise the research and form the basis of the investigation process. The sub questions are:

1. What is open access to research data?
2. Is there a need for open access to research data?
3. What are confronting opportunities and difficulties in the process of open access to research data?
4. To what extent is there a need for a regulatory framework on open access to research data?
5. What are possible types of regulation of open access to research data?

1.3 *Relevance for society and science*

This study is socially relevant because of the actual debate on open access to research data and open data (European Commission, 2012; Rijksoverheid, 2013). The discussion is about to what extent the government needs to regulate the transition and to what extent the European Union comes into play. Furthermore, the developments in the Information and Communication Technology (ICT) and data dependency are expanding which may need regulation and governmental policies (European Commission, 2012; Guibault & Wiebe, 2013). Additionally, the problems of data which are related to privacy and property rights are ambiguous and vague, and research data often lacks legal protection (Guibault, 2013; Guibault & Wiebe, 2013). Therefore, the open access process needs to get full attention of national and supranational governments. The problem rising with the growth of data is a

cross-border problem and a global problem for governments. The Netherlands has suggested to take up a leading role in regulating open access (Rijksoverheid, 2013) and therefore this study is also an important contribution to the debate on data. During the study the relevance for society has been proven several times. First of all, the Dutch Ministry of Economic Affairs mentioned at the beginning of this study that its ministry is working on exact the same questions as discussed in this research (Van Loenen, interview). Additionally, the Dutch ministry of Education, Culture and Science (OCW) has asked for a presentation of the final outcomes of this study. Furthermore, several updates about open data came from one of the Dutch ministries or from Philips during this study. These actions underpin the relevance of open access to research data and its relation to society.

The academic contribution of this study is important, because there is not much known about the consequences for the Dutch academic community. This study contributes firstly to science by adding information on what open access to research data entails and the lessons to be learned in the context of the government and judiciary. Secondly, this study makes contributions to science as a discipline (ontological). This study adds new scientific information, it combines the actual developments with scientific insights using scientific methods of investigating. Furthermore, theory of open access is comprehensive and mainly focuses on open access to scientific publications (Björk, Welling, Laakso et al., 2010; Guibault & Wiebe, 2013; Schmidt & Kuchma, 2012). The results will give more insight in the conditions and overall framework of open access, which is often vague and wide. Structured information about this topic will give insight in future paths of open access.

In addition, this study is a public administration topic, because it gives an advice to the Dutch government on open access to research data. This study discusses a current debate on open data and possible governmental regulation strategies. Public administration is a discipline that deals with the functioning of the society and the governmental agencies. A typical public administration study deals with practical issues that have to do with management, organisation and policies of organisation in the public administration. Open access to research data fits in this context. Furthermore, this study discusses the role of the private sector, the government and other stakeholders. All have different interests and perspectives, this outweigh of interests and approach is typically a public administration topic. This study also describes the social debate and difficulties of open data, what makes this study relevant for a government and is therefore a public administration topic. Additionally, the question if a government needs to regulate is therefore an essential question in the public administration.

1.4 *Outline of the report*

This study has been structured differently than ordinary theses, because of the iterative character of this research. Open access to research data is as mentioned before a new debate and little literature

has been published until so far. Due to the lack of literature, the structure of this study deviates from an ordinary thesis structure. In order to embed literature in this research, a Delphi-analysis has been conducted. This analyzes the current open access process and the opinions about this issue. The Delphi-analysis has been conducted in order to map the contemporary situation of open access to research data and its impact for science and other stakeholders, this will be described in the analysis. Based on this, a theory can be constructed for open access to research data and the way to approach the issue and its methods to process the open access to data. In this way this thesis fits with the current situation of the open access to research data and makes the thesis stronger by using 'first-hand' information and starting at the stakeholders of data. A Delphi-analysis is often used in cases of less (scientific) information and is a method for generating a theory based on the input of experts (Van Dijk & Landsheer, 2011; Hsu & Sandford, 2007; Somerville, 2008; Van Thiel, 2007, p. 110; Okoli & Pawlowski, 2004). This structure is more appropriate for generating new scientific information in the sense of a new theory which could be used for further research.

The structure of this study will therefore be structured as follows. Chapter 1 gave an introduction to this study and the problem definition. Chapter 2 discusses briefly some theoretical background of the open access debate and some general issues related to governmental regulation. The research method will be clarified in chapter 3 as well as the operationalization of key variables. The Delphi method will also be clearly explained in this chapter. The results of the Delphi-analysis are presented in chapter 4. In chapter 5 a theory will be constructed based on the Delphi-analysis and additional interviews. In the final chapter, conclusions will be drawn, the research questions will be answered and recommendations will be given to implement the open access policies smoothly as well as recommendations for further research.

As mentioned before, the next chapter will elaborate the first part of a theoretical framework.

2. Theory | part one

This chapter elaborates the first part of the theory. It entails background information on open access and contemporary European efforts to regulate this. Theory from public administration discusses the regulation possibilities more in-depth. More specifically, this chapter contains general information about the context of the open access debate (2.1). Next, existing legislation on open access to research data (2.2) and the European perspective on open data will be elaborated (2.3). The regulatory focus of this study is essential and therefore more information on regulatory strategies and policy instruments will be discussed (2.4). This chapter ends with a summary (2.5).

2.1 *General context*

After the beginning of the information era around the 1970s, the focus shifted to digital data (Bloem et al., 2013). This so-called ‘big data movement’ refers to the growth of the amount of data particularly digital data (Buchholtz, Bukowski & Śniegocki, 2013; Howe et al., 2008). Digitalisation has spread and extended enormously over the past few years and its effect on society is extensive in the way that people rely more and more on ICT and data (storage). “The volume of available facts is higher than ever before, it expands quicker, comes from many more sources and is expressed in different forms than small and well-structured data from the past” (Buchholtz, Bukowski & Śniegocki, 2013, p. 11). This growth in the amount of data and other ICT developments means that measures are needed for data management and accessibility. Data can be found in divergent sectors and applications in society, from the political process to the assembly of our cell phone devices and storage of our daily email (Manyika et al., 2011). Data are not solely associated with the internet, but became also associated with the global economy in transparency and development (Vu, 2011). Furthermore, the use of ICT in society (ICT penetration; Vu, 2011; International Telecommunications Union, 2014) has positive effects on economic growth by fostering technology diffusion and innovation (Vu, 2011, p. 370). It enhances the quality of decision-making by firms and households and increases demand. Furthermore, it reduces production costs, which together raises the output level (Vu, 2011, p. 370).

The growth of big data and the dependency on ICT requires more and more open access to data (Buchholtz, Bukowski & Śniegocki, 2013). The reliability of scientific results will increase by allowing easier replication of the original analysis. Furthermore, businesses may increase credibility and relations with the public and authorities through opening access to data on their operations (Buchholtz, Bukowski & Śniegocki, 2013, p. 60). The accessibility of data was mainly an academic topic with narrow practical applications, currently it turns to be a “key transformational economic forces of the twenty-first century” (Buchholtz, Bukowski & Śniegocki, 2013, p. 22). Because of large

volumes of digital research data produced by researchers and the digital movement, the importance of storing and sharing such data has increased (Dille & Doorn, 2011, p. 23; Guilault & Wiebe, 2013). **“Open access requires that all the materials need to be not only accessible but also reusable, in terms of the ability to make copies and redistribute them”** (Guilault & Wiebe, 2013, p. 144). The term *open access* was first formally defined at a meeting in Budapest in 2001 (Guilault, 2013; Guilault & Wiebe, 2013). Based on several scientific articles (Canessa & Zennaro, 2008; Guilault & Wiebe, 2013, p. 144), open access to research data includes all original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material. The essential characteristics of open access entail: free accessibility, further distribution, and proper archiving. Open access contributions must satisfy two conditions based on the Berlin Declaration (Canessa & Zennaro, 2008, p. 15; Guilault & Wiebe, 2013, p. 144; Max Planck Gesellschaft, 2003):

- *The author(s) and right holder(s) of such contributions grant(s) to all users a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship (community standards, will continue to provide the mechanism for enforcement of proper attribution and responsible use of the published work, as they do now), as well as the right to make small numbers of printed copies for their personal use.*
- *A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in an appropriate standard electronic format is deposited (and thus published) in at least one online repository using suitable technical standards (such as the Open Archive definitions) that is supported and maintained by an academic institution, scholarly society, government agency, or other well established organisation that seeks to enable open access, unrestricted distribution, inter interoperability, and long-term archiving.*

The Berlin Declaration agreed on the limitation of reproduction and distribution of articles should be to given to authors who control the integrity of their work and the right to be properly acknowledged and cited, because they can be seen as the owner of the research data (Guilault & Wiebe, 2013, p.144; Max Planck Gesellschaft, 2003). However, ownership is a sensitive issue and will be discussed in section 4.5.

Research data in essence provides the evidence for the published amount of scientific knowledge, which is the foundation for all scientific progress (Buchholtz, Bukowski & Śniegocki, 2013; OECD, 2007). Research data are collected, observed or created for the purposes of analysis to produce and validate original research results (Macdonald, n.d.) and includes all kind of data produced in the course of scientific research, such as databases of raw data, tables, graphics, pictures

and so on (Guibault & Wiebe, 2013). The term does not cover laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues or physical objects (e.g. laboratory samples, strains of bacteria and test animals such as mice; OECD, 2007).

In this study research data shall be regarded as data created in a digital form ("born digital") or converted to a digital form (digitised; Macdonald, n.d.). This focus is appropriate because one of the purposes of open access is rapid dissemination of information, which can be best achieved by the internet and digital formats are in this sense necessary. The more data are made openly available, the greater the level of transparency and reproducibility and hence the more efficient the scientific progress becomes (Buchholtz, Bukowski & Śniegocki, 2013; Molloy, 2011). This perspective is becoming mainstream among many funders, publishers, scientists and other stakeholders in research. The demand to more open data solutions is a consequence of the development in ICT and the data-driven economy (Buchholtz, Bukowski & Śniegocki, 2013). More and more depends on scientific data and ICT-driven data.

Open access to research data often relates to research projects that are publicly funded. Information on publicly funded studies and associated challenges are discussed in the analysis chapter (chapter 4). However, defining publicly funded research is essential to fully understand the difficulty of open access to data in certain disciplines. According to the OECD (2007, p. 14), publicly funded research data are: "research data obtained from research conducted by government agencies or departments, or conducted using public funds provided by any level of government". Given the fact that the nature of public funding of research varies significantly from one country to another, the OECD calls for a flexible approach in the case of access to research data. Here, this study focuses on open access to research data in general. The issue itself is difficult enough and in further research the issue of open research data can be more specified.

The Dutch secretary of state Dekker calls for more open access, which focuses on the accessibility of scientific publications (Rijksoverheid, 2013). He proposes regulation of open access in order to make an unitary system of open access in data because of its transnational feature. Partly due to the transnational features, many challenges remain in how to use the massive data sets while ensuring data security, privacy and providing accessibility (Howe et al., 2008; Schmitt, 2013; Tene & Polonetsky, 2012; 2013). Accessibility is the main topic in the data discussion, because data are the backbone of scientific research in the sense that they can give insight in the methods used. However, research data are often not provided, because the scholar's knowledge is economically attractive and needs to be protected in order to keep the authors' rights. When the Guardian started publishing Edward Snowden's leaks about the big data practices of secret services, it underpinned the

discussion about the accessibility and privacy of information (Howe et al, 2008). Therefore, the great privacy debate has to be taken up by organisations and governments in order to regulate full open access. In this context, the European perspective will be given in section 2.3 and the issue of privacy and accessibility will be discussed more in-depth in chapter 4.

2.2 *Legislation*

There is a lack of legislation related to open access to research data (Guibault, 2013). Individual research data and the datasets containing them may not simply fall under the copyright laws, because research data cannot be qualified as protectable subject matter for lack of originality (Guibault, 2013, p. 2). The copyright law does not connect to the original expression of ideas and because of this cannot easily qualify as original. In most of the cases scientific databases do not meet the threshold for copyright protection (Guibault, 2013, p. 20). It can be protected by the *sui generis* database right, protect the collections of scientific data, but then it does need to show: “substantial investment in the obtaining, verification and presentation of the whole or exclusive right to prevent the extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database” (Guibault, 2013). The Netherlands is so far the only member state to have explicitly regulated the implementation of the *sui generis* rights by public bodies. Article 8 of the Dutch Database Act denies a public authority the right to exercise its exclusive database rights unless the right is reserved explicitly by a general statement in an act, order or ordinance, or in a specific case by notification on the database itself or while the database is made available to the public.

2.3 *European perspective*

In this section the contemporary European perceptive on open access will be elucidated. The aim is to give more background information on the process of open access to research data at the European level and some more background information on the issue of open access itself. More information will be given in the analysis (section 4.12).

Prior research shows that open access to research data is essential for the conduct and advancement of science (Arzberger et al., 2004). Improvement of access and sharing of publicly funded research data is an issue that touches on all aspects of the research sector and the development of knowledge, and involves all experts in the conduct of research. This issue has been recognized by the European Union by the European Commission (EC). On July 17, 2012, the EC published a communication which underpins the importance of improving access to research data, which form the basis for the quantitative analysis scientific publications (European Commission, 2012a; Guibault, 2013). The EC calls for coordination and open access definition in member states

(European Commission, 2013a; 2013b). The pace of innovation will accelerate, researchers will collaborate so that duplication of effort will be avoided. Additionally, open access will allow subsequent research to build on previous research results, as it will involve citizens and society in the scientific process (European Commission, 2012; 2013b; Gaubuilt, 2013). A pilot initiative was initiated on open access to peer reviewed research articles in the European *Seventh Research Framework Programme* (FP7), known as the OpenAire project (European Commission, 2012). The EC have taken into account the accessibility issue also in their research and innovation goals until 2020. This *Horizon 2020* program (follow-up of FP7) recognises the issue of open access to publications and scientific research data (European Commission, 2012a). It established an Open Research Data pilot which aims to improve access to scientific information and to boost the benefits of public investment in research. In the next section this pilot will be explained more in-depth.

2.3.1 *Open access to scientific publications*

As mentioned before the open access debate started with the accessibility to scientific data (Björk, Welling, Laakso et al., 2010; Guibault & Wiebe, 2013; Schmidt & Kuchma, 2012). The European Commission defined open access to peer-reviewed publications as the general principle in Horizon 2020 in two possible ways (European Commission, 2012b). Through open access publishing ('Gold' open access) or self-archiving ('Green' open access; (European Commission, 2012b; European IPR Helpdesk, 2014; Rijksoverheid, 2013). The Green open access method means that the published article or the final peer-reviewed manuscript is archived by the author, or a representative, in an online repository before, alongside or after publication (European Commission, 2012b). Access can be given after a period of embargo. The Golden route of open access means that an article is immediately provided in open access mode as published (European Commission, 2012b). The associated costs are shifted away from readers and instead charged to for example the university or research institute to which the researcher is affiliated or to the funding agency support the research (European IPR Helpdesk, 2014, p. 3). In the context of research funding, open access requirements in no way imply an obligation to publish results (European Commission, 2012b). These routes are not mutually exclusive and in projects of the European Union a research consortium can choose the most suitable approach for each publication concerned. The European Cooperation in Science and Technology (COST, 2013) mentions multiple routes of open access, but the green and gold route have been used most often (Rijksoverheid, 2013; European Commission, 2012b).

2.3.2 *Open access to research data*

Besides the access to publication, accessibility to research data have also been placed explicitly on the agenda by the EC. According to the European Commission (2012b), open access to research data

refers to the right to access and re-use digital research data. Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated free of charge for the user. More specifically (European Commission, 2012b):

research data refers to information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview, recordings and images. The focus is on research data that is available in digital form. (p.3)

A communication of the EC (2012b) on open access to research data and the *Model Grant Agreement* (European Commission, 2014) stipulate the open access to research data and are an addition to the pilot. Article 29.3 of the Model Grant Agreement explains the legal requirements for projects participating in this pilot. The areas of Horizon 2020 that participate in the Open Research Data Pilot are specified to seven areas and in appendix C (European Commission, 2013b; European IPR Helpdesk, 2014, p. 9). Other individual projects funded under Horizon 2020 and not covered by the scope of the pilot may participate on a voluntary case-by-case basis.

However, projects may at any stage opt out of the pilot for a variety of reasons, namely if (European Commission, 2013b, p. 9):

- participating in the pilot is incompatible with the Horizon 2020 obligation to protect results if they can reasonably be expected to be commercially or industrially exploited;
- participating in the pilot is incompatible with the need for confidentiality in connection with security issues;
- participating in the pilot is incompatible with existing rules concerning the protection of personal data;
- participating in the pilot would jeopardise the achievement of the main aim of the action;
- the project will not generate or collect any research data; or
- there are other legitimate reasons to not take part in the pilot.

The European Commission requires a Data Management Plan (DMP) from researchers in the open research data pilot (European Commission, 2012b). “The purpose of the DMP is to support the data management life cycle for all data that will be collected, processed or generated by the project” (European Commission, 2013d, p. 3). The DMP should at least entail information about data set reference, description, standard and metadata. Furthermore, information on data sharing and archiving and preservation has been recommended by the EC. The research data should be easily discoverable, accessible, assessable and intelligible. The data should also be useable for other purposes and interoperable to specific quality standards. In the Delphi-analysis the experts will go into more depth on the specifications of a DMP, see therefore chapter 4.

2.4 *Policy instruments*

The European Union has anticipated on the debate of open access to research by establishing a pilot. However, national governments can also play a role in improving access to scientific information by using regulation.

A government does have several instruments at its disposal to regulate behaviour of individuals (Hood & Margetts, 2007, p. 2; Hague & Harrop, 2010, p. 375). “Policy instruments can be defined as a set of techniques by which governmental authorities wield their power in attempting to ensure support and effect social change” (Bemelmans-Videc, 1998, p. 3). In other words, policy instruments are “everything a policy actor may use to obtain certain goals” (Van der Doelen, 1998, p. 131). Various governmental stakeholders are involved in the actual formulation and choice of policy instruments. These stakeholders can be consulted in the phase of policy design or in the implementing phase. Additionally, the choice between two or more alternatives is based on decision making (Bemelmans-Videc, 1998). It can be expected that the design and choice of instruments will vary with the background, roles and cognitive orientations of policymakers. Contextual factors by history also influences the views of instruments. In general, the choice of a policy instrument is based on the following (competing or conflicting) values (Bemelmans-Videc, 1998, p. 7):

1. Effectiveness; which entails the extent that objectives have been achieved;
2. Efficiency; which is the ratio between the input-output and outcome;
3. Legality; is the degree of correspondence of administrative policies with formal rules; and
4. Democracy; is the degree to which administrative policies correspond with accepted norms.

2.4.1 *Typologies*

Policy instruments can be categorized in the purpose or role of policy instruments (detectors and effectors), the types of instrument (sticks, carrots and sermons) and the goal of policy instruments (information, facilitation, incentive and regulation).

Policy instruments can be distinguished along their purposes for detection and for effecting. “Detectors are all the tools government uses for taking information” and “effectors are all the tools government can use to try to make an impact on the world outside” (Hood & Margetts, 2007, p. 3). Effectors are more about influencing society, which is more appropriate in this study. Therefore effectors are elaborated in this section.

According to Vedung (1998), policy instruments can be classified in three types regulation, (sticks), economic means (carrots) and information (sermons; Hague & Harrop, 2010, p. 375). This classification has been illustrated in figure 1. According to Hague & Harrop (2010), a list of policy instruments can be made, but most policies use a combination of tools (Peters & Van Nispen, 1998).

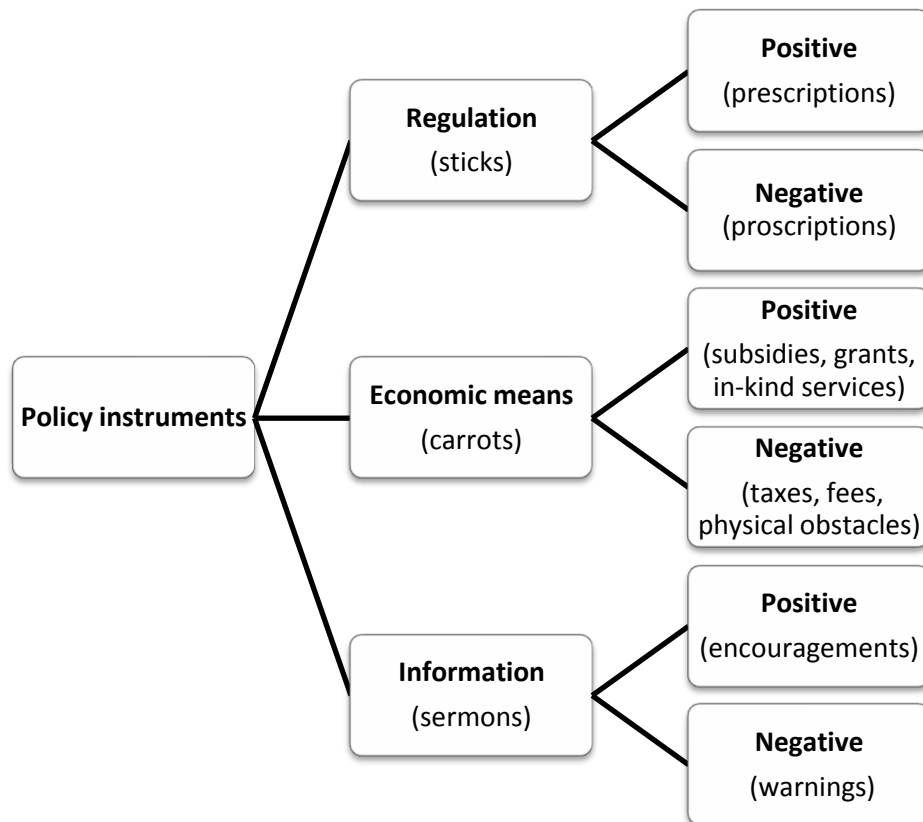


Figure 1. Policy instruments. From *Carrots, sticks and sermons: policy instruments and their evaluation* (p. 250), by M.L. Bemelmans-Videc, R.C. Rist and E. O. Vedung, 1998, New Brunswick, NJ: Transaction Publishers.

Firstly, regulation, this can be defined as measures taken by governmental units to influence people by means of formulated rules (Vedung, 1998, p. 10). Regulatory instruments are “used to define norms, acceptable behavior or to limit activities in a given society” (Lemaire, 1998, p. 59). This regulation measures, also known as the government’s stick entail traditional command and control functions, as for example banning and requiring certain issues. The law and a certain sanction are the stick for prescribing or preventing certain behaviour. These sticks can be affirmative by using prescriptions or negative by using proscriptions (Vedung, 1998). For example in the case of health and safety issues, governments want to give a clear signal to society of no compromise and strong action. Therefore, regulatory instruments can be more appropriate than other instruments, which will be described hereinafter. In order to determine if regulation should be established or amended, governmental agencies should demonstrate that there is a problem or risks and that the benefits of regulation outweigh the costs (Lemaire, 1998, p. 66).

Secondly, economic means, that is providing or taking away of material resources, this can be in cash or in kind (Vedung, 1998). Economic instruments make certain behaviour cheaper or more expensive in terms of money, time, effort and other valuables. In some cases, addressees may decide not to make use of a government incentive (grant) or disincentive (taxes on tobacco). Economic

instruments, also known as carrots, can be affirmative by using subsidies, grants or in-kind services, for example free medical services (Vendung, 1998). Also carrots can be negative because it is an extra obligation in the payment of buying for example tobacco. Taxes, fees or physical obstacles fall under this category.

Finally, information instruments, known as sermons, can be distinguished. Sermons contain all the efforts seeking to demonstrate the concern of the government (Hague & Harrop, 2010). This instrument attempts at influencing people through the transfer of knowledge (Vendung, 1998). It is a 'catch-all' term for all communication campaigns. Examples are brochures, folders, commercials, inspections, training programmes or educational efforts (Vendung, 1998, p. 33). The government offers data, facts, knowledge, arguments and moral appeals by informing citizens about what is good or bad, right or wrong (via for example mass-media). Moreover, it can "provide information about what people are allowed to do or how they should act and behave" (Vendung & Van der Doelen, 1998, p. 104). Therefore, it can be positive (encouragements) or negative (warnings) information.

Also other typologies can be made according to different scholars, a common fourfold typology will also be discussed (Balch, 1980; Peters & Van Nispen, 1998, p. 17; Vendung 1998). This typology overlaps the aforementioned typology by several aspects or even the same name of strategy. However, this typology emphasizes specific aspects of policy instruments. This typology focuses more on the goal of regulation compared to the stick, carrots and sermons typology. Therefore, this typology will be discussed separately. The classification of sticks, carrots and sermons will be applied to this typology, because it is often used in regulation literature (Peters & Van Nispen; Bemelmans-Videc, Rist & Vedung, 1998). The fourfold typology has been illustrated in figure 2.

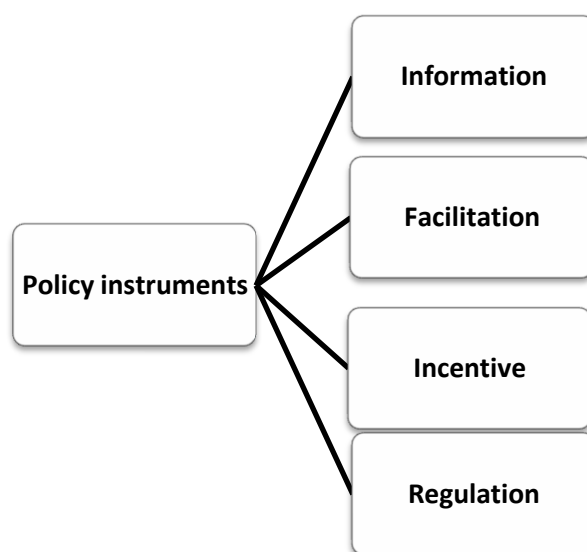


Figure 2. Policy instruments. From "The stick, the carrot, and other strategies: A theoretical analysis of governmental intervention", by G. I. Balch, 1980, *Law & Policy*, 2(1), 35-60.

Four policy instruments are proposed by Balch (1980). Firstly, a government can inform target individuals of benefits (Balch, 1980, p. 35). This is the so called sermons classification as mentioned before (Hague & Harrop, 2010; Raadschelders, 2003, p. 269). Often people fail to use a service or behaviour because they are unaware of it or uncertain about its consequences. On the other hand, many of the individuals or organisations can be aware of the benefits, but do not know how to obtain or use them. In these cases there is a gap of information between potential and new behaviour and information should therefore be provided (Balch, 1980, p. 35). Information should be clear and relevant, because it enhances the likelihood of attention, favourable perception and persuasion of certain information (Balch, 1980, p. 43). Additionally, appropriate incentives are essential in the provision of information. “Information tightens the link between behaviour and reinforcer”, which increases also the benefits to the individual (Balch, 1980, p. 43). Furthermore, information can ‘add value’ to a product, service or behaviour by opening a new market (Balch, 1980). The amount of given information is ambiguous, because creating and spreading information can change the competitive situation of for example the industry (Balch, 1980). Especially in change cases the ‘added value becomes clear, because providing information motivates change and may become more legitimate (Balch, 1980, p. 43). In the case of the government, the provision of information should be between vague and detailed. A proper guideline cannot be given (Balch, 1980). The information needs at least be uniform, but also applicable to different people and situations.

Secondly, *facilitation* strategies may change the behaviour of individuals (Balch, 1980, p. 37). According to the classification of Hague and Harrop (2010), this strategy could also be called as the carrot instrument (Raadschelders, 2010). The adaption of new behaviour will get easier by setting up facilitation strategies, because new behaviours, products or services may take expenses, skills or time. Factors that may constrain behaviour or services must be removed in order to raise the likelihood of its occurrence (Balch, 1980). Obstacles as changeover costs or specifically required facilities. Governmental assistance may overcome those obstacles in order to achieve the desired activity. Grants may facilitate behavioural change through reinforcement processes (Balch, 1980; Hague & Harrop, 2010, p. 376). Raadschelders (2003, p. 269) defines this facilitation strategy as an economic instrument. Negative measures can be taken, for example fee for using a specific road. Positive measures could be subsidies, grants or contracting-out et cetera. However, programmes or behaviour may disappear when grants are withdrawn, because of the fixed, non-contingent method of reinforcement (Balch, 1980, p. 47). Too much facilitation could induce opposition to governmental interference in affairs (Balch, 1980). Facilitation itself can be encouraged by giving people a chance to participate in the choice and implementation of modifying their own behaviour (Balch, 1980, p. 47). In this way desired output or compliance of certain behaviour can be increased. On the other hand too much freedom can be a barrier, instead of being a facilitator itself.

Thirdly, *incentive* strategies can be a method for changing behaviour (Balch, 1980). This strategy has a non-coercive character. This strategy leaves maximum discretion to users and producers, because producers and consumers work through the market to reach the combination of supply and demand which best satisfies their mutual interests. This mechanism could be seen as an incentive (Balch, 1980). A government in this sense “provides no facility, but merely reduces the cost of the desired activity, or increases the cost of its competitor, or both” (Balch, 1980, p. 38). Other strategies by comparison have different approaches in affecting behaviour. For instance, the information strategy reduces the costs of obtaining information while facilitation strategies make adoption more easy. Regulatory strategies threaten negative benefits when individuals fail to perform expected action. Compared to the incentive strategy, this could be seen as a more nuanced or mixed way of affecting behaviour or activities. Incentive strategies may affect behaviour negatively or positively by stimulating or not stimulating certain output or behaviour (Balch, 1980). Examples of strategies with a negative effect are taxes and insurance benefits. Positive reinforcements are transfer payments and discounts (Balch, 1980, p. 52). Incentive strategies get people to change their behaviour by raising the relative attractiveness of the desired behaviour and the beneficiaries changes routines or activities voluntarily.

Finally, *regulation* strategies can be used to change the behaviour of individuals and is legal in nature (Baas, 1995; Balch, 1980; Peters & Van Nispen, 1998). Balch (1980, p. 48) relates a regulation strategies to punishment or threat of punishment as a method for changing the behaviour and is therefore a stick in the sense of Lemaire (1998). The regulation instruments has an instrumental function and it also possess a normalizing and guaranteeing function (Peters & Van Nispen, 1998). Punishment is often used to reduce a specific behaviour, not to increase or stimulate certain activities or behaviour. Specifically, it is hard to punish someone for not using some devices or performing some activities (Balch, 1980, p. 48). Punishment as said before can suppress behaviour. Recent studies have concluded that the rate of compliance with laws is associated with the certainty, quickness and severity of punishment for non-compliance (Balch, 1980). An effective punishment must be repeated often and extensive surveillance or monitoring of behaviour is also necessary, according to Balch (1980, p. 49). Furthermore, regulation needs an agency to do the regulation which also needs to take up a watchdog function. In order to check the effectiveness of the regulation and its implementation. Additionally, punishment works best when other factors are appropriately manipulated, especially in cases of positive reinforcement (Balch, 1980, p. 49). Unforeseen circumstances need to be reduced, removed or replaced.

However, punishments can actually facilitate the target behaviour, but can also be counterproductive (Balch, 1980, p. 50). First of all, it may elicit emotional effects that stop the punished behaviour because they interfere with it (Balch, 1980, p. 50). For example, anger and

anxiety. Furthermore, an effect of punishment on certain behaviour might be a fighting response which is also counterproductive. For example, the necessity of a buzzer in a car when the seatbelt is not fastened. Drivers are therefore forced to follow the rules by fastening the seatbelt. However, it can be avoided by disconnecting the buzzer. This fighting response can also occur and does not improve the compliance. All in all, “regulation may be appropriate when there are few ways of achieving the desired result and when the desired result is specific and easily calculated” (Balch, 1980, p. 51).

In general, Information and facilitation work best on ‘motivated’ people. “Regulation works best for strongly desired, discrete, detectable goals achievable in few ways; but it requires much monitoring and may promote undesired reactions” when (Balch, 1980, p. 35). Incentives are the basis of the most reliable, efficient strategies, especially if the change is continuous, detectable, and achievable in many ways.

2.4.2. *Hybrid regulation*

“The effectiveness of state interventions cannot be separated from their legitimacy”; governments should act one way and not hesitate to implement the other way as well (Van der Doelen, 1998, p. 129). So, the approach and style of the government is essential for striving for certain behaviour or output. Legitimacy can be increased by using hybrid regulation (Levi-Faur, 2011; Black, 2002). In general, hybrid regulation models are combinations of governmental and non-governmental agencies or combinations of several layers of governmental agencies (Black, 2002; Levi-Faur, 2011). Hybrid regulation models can be seen as the new policy instruments in current society (Peters & Van Nispen, 1998; Jordan, Wurzel & Zito, 2005). It can shortly be divided in four types of hybrids (Levi-Faur, 2011, p. 10) and has been illustrated in figure 3. Firstly, co-regulation, in this case the responsibility of the regulatory enforcement is shared by the regulator (government) and regulatees (stakeholders in society; Levi-Faur, 2011). An example is the Dutch railways by which the government gives a framework, but the implementation has to be done by a stakeholder in the private sector.

Secondly, self-regulation (Hague & Harrop, 2010, p. 377; Levi-Faur, 2011). The regulator forces the regulatee to write a set of rules that fits a set of contingencies in that organisation. In this way the regulatee needs to enforce the rules instead of the government. This type of hybrid regulation is appropriate in situations where a goal will not be achieved through prescriptive rules imposed by rulers rather by **encouraging the aim of the regulation**. However, regulation instruments are highly coercive and are sanctioned by negative or affirmative measures (Bemelmans-Vedic & Vedung 1998, p. 250 in Raadschelders, 2003, p. 269).

Thirdly, meta-regulation by which government regulates its own regulation (Levi-Faur, 2011; Raadschelders, 2003). In a broader sense it means that any form of regulation (for example law or tools) that regulates any other form of regulation (Parker, 2007).

Fourthly, multi-level regulation (Levi-Faur, 2011). The regulatory authority is allocated to different levels: supranational, national, regional or local. The authority can be allocated based on a functional basis, a hierarchical basis or just a result of an incremental, path-dependent process. In the case of a functional basis the regulatory authority is allocated to the layer what can handle the issue based on capacity (Levi-Faur, 2011). The hierarchical basis works on the basis of a supreme authority in each layer. In the case of an incremental process, the multi-level regulation and its coordination is a result of the amalgamation of different agencies, what happened slowly over time (Levi-Faur, 2011).

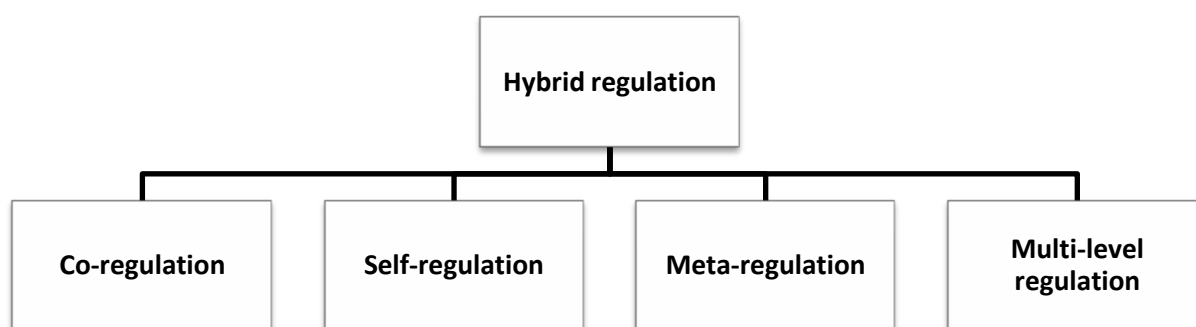


Figure 3. Hybrid regulation. From *Handbook on the Politics of Regulation* (p.10), by D. Levi-Faur, 2011, United Kingdom: Edward Elgar Publishing.

2.4.3 Digital era

In current digital era the challenge for governments is **to find new ways of using a limited basic array of instruments effectively and creatively** as technology and social patterns change (Hood & Margetts, 2007). Governments should use its ‘nodality’ position in a sharper and smarter way than before (Dunleavy & Margetts, 2000; Hood & Margetts, 2007). **Nodality** is “the property of being in the middle of a social network” (Hood & Margetts, 2007, p. 21). It is the ability to receive information from societal actors and to transmit messages which are accorded special attention by them (Dunleavy & Margetts, 2000). The nodality tool can be seen as a new tool compared to the aforementioned classical tools. Additionally, it is timeless and a cheap tool for governments now and in the future (Hood, 2007; Dunleavy & Margetts, 2000).

In particular, “the greater a government’s nodality, the more likely that it can use the dissemination of information alone to change societal behaviour”(Escher, Margetts, Petricek & Cox,

2006, p. 4). If the government does not keep up with current trends in the digital era, then government's nodality position will decrease (Dunleavy & Margetts, 2000). Its positions can be undermined if public agencies become relatively invisible on the internet and email networks compared with private sector and civil society organisations (Dunleavy & Margetts, 2000, p. 18). This competition for nodality increase because of the differences across and within governmental and commercial sectors (Escher, Margetts, Petricek & Cox, 2006). For example, websites – as a tool for nodality – vary in the extent to which they are accessible, visible and connected to other parts of the internet. Where a government faces greater competition for nodality, it may need to be more active in its effort to obtain information or put its messages across (Hood & Margetts, 2007, p. 197).

According to Hood and Margetts (2007), government should have a smart approach by informing citizens appropriately. Firstly, governments could ensure that its own websites figure in the top ten or twenty sites listed when a search query is made through a search engine. This increases the visibility of the government and the government can take up its role as main communicator of information. Additionally, internet users should find the information they need quickly on websites. This also increases the usability and this may increase indirectly the visibility of governmental communications. Besides on this, scalability, “the ability to apply effectors at a variable range of intensity rather than in on/off mode” (Hood & Margetts, 2007, p. 200), is also something that could help the government to improve its position in a digital era. In this way government can hit the target just as hard as it needs to be hit, rather than always an equal approach of each governmental communication. Another aspect of the digital era and the nodality position of a government is that policy instruments need to be direct (Hood & Margetts, 2007). Directness is “the precisions with which an instrument can be directed to a specific beneficiary or maleficiary” (Hood & Margetts, 2007, p. 158). However, directness is difficult in situations where governments use a general instruments to influence the behaviour of an individual or a specific group. Even indirectness can be effective, but often externalities diminish the effect of such a indirect tool.

Summarising the aforementioned, regulation of a government in a digital age needs to be **active, informative, focused** and to some extent **flexible**.

2.4 *Summary*

The growth of data and the digitalisation leads to more open access to research data. Research data in essence provide the evidence for the published amount of scientific knowledge. Open access gives opportunities for society, however several challenges, including lacking legislation, do exist. The debate on open access to research data is new and therefore needs more understanding and clarity. Therefore the European Union, as supranational government, is working on this theme and established a pilot in the Horizon 2020 programme which is a trail of open access of research data in

practice. National governments also play a role in the interpretation of open access to research data. Several strategies, such as regulation and incentives, may change the behaviour of individuals. Hybrid regulation, such as co-regulation, is also very important in regulating issues. In general, a government should be active, informative, focused and flexible.

This chapter has offered some first insights into this new phenomenon. Next, more empirical insights have been collected, that will be used to further elaborate the theory.

3. Research methodology

In this chapter the methodological framework will be discussed. This is the basis and backbone of the research. Firstly, a short notification will be presented (3.1), followed by sensitizing concepts (3.2), using the selected research methods (3.3). The validity and reliability will also be discussed (3.4). Finally, a number of limitations will be discussed, which are indirect recommendations for future research (3.5). The chapter concludes with a summary (3.6).

3.1 *Notification*

Open access to research data is a new debate and little literature has been published until so far. Therefore this study has been structured differently than ordinary theses. By using a different structure, the topic can be approached bottom-up. Experts will give input as basis for a regulatory framework via a Delphi method and interviews will give additional in-depth information about open access to research data. Documents serve as additional information.

The study started with basic information on what the open access movement is. Regulatory theories have been discussed. This information is the first input for the analysis. After the analysis, the second part of the theory will be presented. This is the application of the first part of the theory on the input from the analysis. Combining both parts of the theory and the analysis (Delphi method, interviews and documents) will give an answer to the research question.

3.2 *Sensitizing concepts*

This topic does have an iterative character and therefore an operationalization cannot be made. Sensitizing concepts are more appropriate in this sense, because these concepts function as guideline for the Delphi method and the analysis (Boeije, 2005; Bowen, 2006; Charmaz, 2006). A sensitizing concept is a starting point of a researcher on the way to look at its study, especially in cases if the further research process is unclear (Van den Hoonaard, 1997). Elements of theory provide sensitizing concepts for the Delphi-analysis. Furthermore, the concepts provide a theoretical foundation for the development of research (Bowen, 2006). "Sensitizing concepts give researcher initial but tentative ideas to pursue and questions to raise about their topics" (Charmaz, 2006, p. 30). It functions as a point of departure for studying empirical data and could be helpful to code data.

Sensitizing concepts in this study are: research data, stakeholders, difficulties, opportunities, regulatory framework and policy instruments (information, facilitation, incentive and regulation). Other sensitizing concepts are: sticks, carrots, sermons, data management, privacy and security. According to these concepts the topic lists for the Delphi-analysis and interviews have been compiled. During the research several (more in-depth) concepts have been added in consecutive

rounds of Delphi-analysis, because the knowledge about the topic increased. Examples are: incentives, research community, licensing and authenticity.

3.3 *Research method*

The open access 'movement' is an issue which may use more understanding and this qualitative research is helpful in understanding the context of the open access to research data, because it gives in-depth information. As mentioned before, open access to research data is a new topic and little literature have been published. In order to receive as much as possible input for establishing a regulatory framework a qualitative research approach has been chosen. The starting point of this study is to define the social environment in order to understand the constructions and experiences in society (Boeije, 2005). Additionally, qualitative research is helpful in situations with limited academic literature (Boeije, 2005). Another reason for using a qualitative approach is that the process of open access to data is complex and variable, an open method is therefore appropriate (Boeije, 2005, p. 36). Moreover, according to Boeije (2005, p. 36) a qualitative research method is appropriate in situations of exploring processes and experiences, which is the case in this study. Open access to research data is a new debate and a qualitative approach will therefore give the most and in-depth information. Especially, the way of forming a possible governmental framework for open access to research data is unknown and a qualitative research is therefore also most appropriate. Additionally, this study aims to describe and map the open access to research data process and aims to offer recommendations (Boeije, 2005, p. 27).

According to the definition of a qualitative research given in the 'Handbook of Qualitative Research' (Haafkens, 1997 in Boeij, 2005, p. 26): "qualitative research is multi-method in focus, involving an interpretive, naturalistic approach to its subject matter". Therefore, several research methods have been used in this study also in order to perceive several perspectives on the open access issue (Boeije, 2005).

Methods to be used include a content analysis of documents, a Delphi-analysis and interviews. This triangulation increases the validity (see section 3.4). Firstly, an extended literature search was carried out to collect (background) information about and experiences with open access to research data. The data sources are official documents and academic literature. The amount of official documents is higher, because there is not that much academic literature yet.

Secondly, a Delphi-analysis is used as the main source for new information (Hsu & Sandford, 2007). The Delphi method is a method to map opinions of experts and stakeholders in order to create consensus (if possible) about complex subjects (Van Dijk & Landsheer, 2011; Hsu & Sandford, 2007; Somerville, 2008; Van Thiel, 2007, p. 110; Okoli & Pawlowski, 2004). More specifically, the

method is based on iteration with controlled feedback (Dalkey, Brown & Cochran, 1969), and characterized by anonymity, information gathering in several rounds and structured feedback processes which ultimately leads to a list of sound information (Hsu & Sandford, 2007; Somerville, 2008; Okoli & Pawlowski, 2004). The Delphi technique normally consists of three or four rounds, involving a series of questionnaires, each building on the results of the previous one (Somerville, 2008). The choice for three information rounds in this study is because of the limited research time. Furthermore, three information rounds are sufficient for answering the research questions and to cover the issue of open access to research data. Moreover, most changes in participants' responses occurs within the first two round and not that much is gained in further iterations and rounds (Mitchell, 1991 in Somerville, 2007), so information will be gathered efficiently in three rounds. A Delphi-analysis is also highly usable in situations of defining the pros and cons associated with potential policy options (Linstone & Turoff, 2002) and is therefore usable by exploring whether governmental intervention is needed to stimulate the open access process. In section 3.2.1 more in-depth information about the Delphi method has been given.

The advantages of the method are that information can be gathered from a geographically diverse panel of experts (Somerville, 2008, p. 2), which is highly necessary and applicable in this study of open access to research data in Netherlands, by using experiences from other countries and best practices. Furthermore, the Delphi method is highly applicable in situations where there is less (scientific) knowledge and the information sought is informed judgement, as it is in this study (Somerville, 2008). Additionally, in cases of uncertainty on both the nature of the problem and the possible policy measures a Delphi method can give new (founded) opinions from experts (Adler & Ziglio, 1996, p. 21). Moreover, a Delphi method is also appropriate in cases where there is no existing information on the size of the problem and experts have knowledge from the field about a certain issue. For example, experts may explain unknown correlations, opportunities or difficulties. The Delphi-analysis is also set up as group communication process in order to achieve convergence on a specific issue (Hsu & Sandford, 2007). Another aim of a Delphi-analysis is to correlate informed judgements from a wide range of disciplines. This study about open access to research data is a topic which is broad and relates to several disciplines such as the legal, scientific, administrative and the social and private spheres, so the Delphi-analysis is appropriate in this context.

Thirdly, next to the Delphi-analysis several interviews have been conducted. First of all, orientating interviews before and during the Delphi-analysis, which gave more background information and served as starting point for a better focus of the theory and application of the theory. These interviews are part of the analysis and additional to the Delphi-analysis. Verifying interviews have been conducted and are complementary, because of the search for more in-depth information on open access to research data. The open access debate is an issues which is not clear

and continuing the dialogue is therefore necessary. Interviews give more accurate information and there is room for following up questions (Boeije, 2005). Seven interviews were held with people who are closely associated with the issue of open access. The chosen interviewees consist of governmental administrators at the national and European level, an expert from the business and experts in the science. The experts mainly come from the Netherlands, which gives more insight in the opinion of Dutch stakeholders and is therefore applicable for giving recommendations for the Dutch government. More information about the interviewees and their professions can be found in appendix B. The interviews are semi-structured in the sense that prepared questions were asked but the interaction with the interviewee also structured the focus of the interview (Boeije, 2005, p. 57). The questions which have been used during the interviews are based on the input of the Delphi-analysis and the given framework. The analysis of the interviews have been done by recording the conversations and writing the questions and answers on paper. After processing the new input from the interviews, information that fits into the analysis chapter have been added by explicitly mentioning the interviewee. For example, given information about a framework on open access to research data during an interview has been written down in the analysis chapter.

All in all this mix of methods increases the scientific foundation of this study.

3.3.1 *Delphi in-depth*

As mentioned before, the Delphi-analysis consist of three rounds. In figure 4 the Delphi method has been illustrated, including input arrows and review/feedback arrows. In appendix I the Delphi-analysis questions have been presented. The first phase is characterized by exploring the subject, wherein each experts contributes his/her opinion and information by answering given questions (Linstone & Turoff, 2002). The open-ended questionnaire serves as the basis for requesting specific (general) information about the subject (Hsu & Sandford, 2007). After receiving the input in the first round, the information needs to well-structured by coding the answers into clusters. After compiling the information, it has to be returned to the experts (Somerville, 2008). General questions are appropriate in this contexts: What is open access? Is there a need for governmental intervention? In appendix H the questions of this Delphi-analysis of open access to research data are included.

In the second phase all participants receive a second questionnaire and are asked to review the summarized information from the first round (Hsu & Sandford, 2007). If necessary, the participants may rank the given answers by prioritizing them. The second round will be result in areas of disagreement and agreement (Hsu & Sandford, 2007).

In the third phase each participant receives the final questionnaire that includes the items and ratings summarized from the previous rounds (Hsu & Sandford, 2007). Participants can give their last revision of the conclusions and opinions or may specify the given answers. This third phase is the

concluding phase, where underlying reasons can be brought to the front and where last evaluations can be given (Dalkey, Brown & Cochran, 1969). With these final given reflections on the questions and remarks from the first and second rounds a conclusion can be drawn.

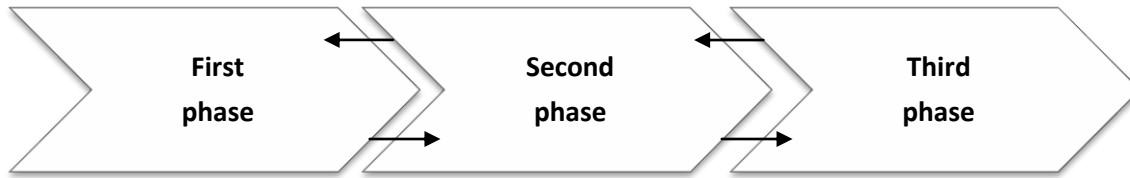


Figure 4. Delphi method (based on Hsu & Sandford, 2007; Linstone & Turoff, 2002).

Several aspects about the Delphi method are insightful to mention. First of all, the period to respond to each round is between one and two weeks, which is adequate according to Hsu and Sanford (2007, p. 4). In their study they recommended two weeks for answering the questions.

Another aspect are the experts in this study. The experts in the Delphi-analysis are chosen for their expertise on open access to research data (Somerville, 2007) and are selected from the existing network of Neth-ER, or have been proposed by researchers from universities, ministries and research organisations. These experts cover a wide scope of opinions. In total ten experts participated in the Delphi-analysis in this study, which is in line with the academic literature. A group of six to twelve participants has been determined as optimum to come to convergence (Hogarth, 1978; Mitchell, 1991). Others recommend ten to eighteen participants (Okoli & Pawlowski, 2004). Although, there is no consistent relationship between panel size and effectiveness criteria for arriving at consensus among experts (Rowe & Wright, 1999; Okoli & Pawlowski, 2004). Additionally, the participants have been chosen because of their experience in the context of universities, research or policies. However, during the process the identity of experts have been made anonymous. This is essential to allow a sharing of responsibility that is refreshing and without inhibitions of experts (Dalkey, Brown & Cochran, 1969, p. 17). Experts can adapt their answer to other experts if they have the feeling that others are more knowledgeable.

The Delphi-analysis has been conducted by online questionnaires sent by email. Online questionnaires are preferable, because of the geographical diversity of the panel of participants and the limited time for conducting this study (Somerville, 2008). A Delphi-analysis can involve qualitative and quantitative data (Hsu & Sandford, 2007). The data will be coded by placing together the same characteristics and (to some extent same) answers. The analysis is structured in a document where each question and answer have been positioned. The answers of participants have to be given in a certain document format which has been attached in the email. Patterns in answers (for example starting each question with the same experts and her/his answer) could harm the

anonymity, therefore all the answers have been positioned without a systematic order. Furthermore, a certain ratio will be given by mentioning how many experts have given the same answer, this will identify convergence because consensus on a topic can be decided if a certain part of the answers falls within the same code (Hsu & Sandford, 2007, p. 4). Answers given by only one expert will be mentioned too, because not all of the experts will come up with the same answers.

3.4 *Research quality*

The quality of the study mainly relates to the reliability and validity of the research method. Reliability refers to whether the study is disturbed by accidental or unsystematic errors (Boeije, 2005, p. 145; Universiteit Leiden, n.d.). Whether and to what extent random errors occur, can be determined by repetition. If random errors occurs, repetition gives a different results than the original study or research. If there are no random errors, the same result will occur by repeating a research. In general, by repeating the questions and the given input the reliability increases. The reliability of this study could be qualified as moderate, because of several aspects. First of all, the Delphi-analysis increases the reliability to some extent because of its accuracy (Boeije, 2005; Van Thiel, 2007). By asking questions in several rounds and using the input from the previous round reliability increases, because it measures the same phenomenon. To a certain extent it also is an extra test of the information given in a previous round by participating experts. By using this method, the information can be refined and updated until consensus has been reached. If there is no consensus, then this will explicitly be mentioned by giving the different opinions. However, the reliability in this study is to some extent limited, because the analysis is mainly based on interviews and the Delphi-analysis. By repeating the whole study by asking the same questions to experts in the Delphi method and during interviews may give a different answers because answers can differ in time. Although, the moral and scope of the answers remain the same, since - for example - the problems and opportunities will be mentioned. Furthermore, by submitting the given input by experts to the same experts will function as a control mechanism which influences the reliability in a more positive way.

The validity is the extent to which the study measures what it should measure (Boeije, 2005; Universiteit Leiden, n.d.; Van Thiel, 2007). In other words, to what extent the observations (methods) correctly answer the research question. The validity of this study is high at some aspects, but several negative aspects play also a role in this study which decreases the validity. One of the first issues is that the participation of interviewees and mainly the participants in the Delphi-analysis is limited (Boeije, 2005, p. 145). Not all of the stakeholders in the debate of open access to research data are involved in this research, which negatively affects the research validity. Additionally, problem of double hermeneutics can occur, because the Delphi-analysis input can be different interpreted which

may give a (biased) different focus on the study (Ginev, 1998; Van Thiel, 2007). However, the individual responses and the interactive interviews affect the validity positively and may correct the different interpretations. This member check can be helpful, because it may open new issues and debates (Boeije, 2005; Van Thiel, 2007). Experts may fully give their (deviating) opinion and additional comments on the several rounds of the Delphi-analysis, which increases the validity of the study. The Delphi-method generates new theory and is therefore appropriate for answering the research question of this study (Boeije, 2005). This affects the validity positively.

The external validity of the results is lower, because the information used in this case is difficult to apply to other situations (Boeije, 2005). The information about open access to research data is highly specific and difficult to use in other situations. Open access to scientific publications for example is already a different case with different circumstances.

The quality of this study is also positively affected by triangulation. Triangulation of methods has been used by using several research methods; data analysis, Delphi-analysis and interviews (Boeije, 2005, p. 152). The 'information feedback' to experts by sending the summary of the interviews increases the quality of the study (Boeije, 2005, p. 153). In this way experts can check if the information corresponds with their knowledge and expertise. If necessary, experts may add information to increase the quality.

3.5 *Limitations*

The first limitation of this study is the lack of (scientific) literature about open access to research data. The issue is upcoming and less awareness exists in society about open access, especially related to research data. Issues related or similar to open access do not exist, so related literature could not be found. Open access to research data is not often investigated, so literature about experiences do not exist and no experiences from other (European) countries were found. To counter this limitation, the structure of this thesis has been changed in order to anticipate on the lack of information. The research method, a Delphi-analysis, is also anticipating on the lack of scientific data by establishing a theory via experts.

The second limitation is related to the aforementioned, because of the lack of literature and awareness the amount of experts is limited. The issue is highly complex, because of the many stakeholders, their interests and the lack of experience and knowledge. Several institutions are working on the issue of open access to research data, but this is not equally divided over the (European) countries. Because of the limited amount of experts, the input could be one-sided and less profound as necessary for a policy framework. As mentioned before, not all the stakeholders are involved in the Delphi-analysis which also affects the outcome negatively by probably sketching a

biased vision on the demands for a framework for open access to research data. Information about open access in other countries is not available, probably because of the lack of awareness. Furthermore, the perspective from publishers has not been described due to lack of time. The lack of experts and information may give an one-sided picture of the open access to research data. These issues have been recognized by explicitly mentioning in the conclusion that it does not give an overall perspective and more research is necessary in order to fully embed the open access to research data. The recommendations also give more insight in further research.

A third limitation is the input of participants in the Delphi-analysis. The input from some of the participants is incomplete, because some of the participants did not fully fill out the questionnaire. Furthermore, some of the participants gave brief answers to the questions. These circumstances may affect the outcome of the analysis. In order to handle this limitation, the final conclusion of the Delphi-analysis has been verified by an interview with someone who has got an overview on open access to research data, because he is working at an European level.

The fourth limitation is the possibility of misunderstanding the questions by the participants (experts and interviewees) and their answers by the researcher. The latter is also called the problem of double hermeneutics (Ginev, 2000; Van Thiel, 2007). The experts could interpret the questions in the Delphi-analysis differently and respond in a different way as intended. The answers were sometimes short. This affected the foundation of the Delphi-analysis. Additionally, interviewees can misunderstand questions during an interview. However, these limitations have been marginalized by presenting each participant in the Delphi-analysis a combination of the same answers. 'Irregular' data (for example one completely different answer than the other participants) may assume a wrong interpretation of the question. By mentioning the combined answer and the irregular answer, experts could comment on that information by adding more information or a different focus. In general, participants could respond to the input of a previous session which functions as peer review, which identifies also irregular answers and strengthens the quality of answers. Additionally, the researcher need to repeat his/her question if the interviewee does not (fully) answer the question.

A fifth limitation is the complexity of open access to research data. The issue is such a new debate in society and for me as a researcher that it took me a lot of time to get familiar with the main issues in the open access to data debate. Furthermore, the structure of an ordinary master study had be changed because of the lack of scientific literature and information. This affected the research progress also negatively. However, as a researcher I indulged in the available information on open access to research data and spoke to experts who gave more background knowledge. The Delphi-analysis process also gave more knowledge about open data, so learning by doing was a way of getting familiar with the topic.

The final limitation are time constraints. Because of the lack of time, several decisions had to be taken to achieve the deadline. As mentioned before, not all of the stakeholders could be involved in the Delphi-analysis and the other (European) countries could not be fully consulted about their experiences. Furthermore, not all of the related issues could fully be discussed, examples are the legislation and further details of a policy framework. The latter is also because of the lack of experience of implementing open access to research data by experts. Specifications are therefore difficult to establish as expert and as researcher. By giving focus and not including all aspects, the thesis could be completed in limited time. By giving recommendations at the end of this study, this study also functions as a starting point for future research.

3.6 *Summary*

This qualitative and iterative study is aimed at mapping the situation of open access to research data. Triangulation by literature analysis, a Delphi-analysis and interviews affect the quality of the study positively. However, several limitations are mentioned which reduce the quality. However, these limitations are tackled by several measures. The reliability of this study results is qualified as moderate and the validity of this study as high.

The following chapter presents the results of the Delphi-analysis, several interviews and the document analysis.

4. Analysis

In this chapter the analysis of the results of the Delphi study will be presented. Additionally, information from the interviews will be given in this chapter. Based on this analysis the theory can be further developed in chapter 5. This chapter starts with a notification before reading the other sections (4.1). The following section starts with the question what open access to research data exactly is (4.2), including information about research data definitions, sensitive data and commercial data. The question whether open access is a solution to a problem will be discussed (4.3) and the necessity of open access to research data (4.4). An important issue is ownership of data, which will also be presented (4.5). Followed by the opportunities (4.6) and the difficulties of open data (4.7). The major question whether there is a need for intervention will be presented (4.8), followed by criteria for a regulatory framework (4.9). In this section issues as data management, authenticity, storage, maintenance and licensing will be discussed. Furthermore, protection of data and the feasibility of the regulatory framework will be presented. The importance of incentives for researchers will be discussed (4.10). Followed by the perspective of the businesses and industry, in particular Philips (4.11). This chapter concludes with Information about other countries and more in-depth information about the European approach of open access to research data (4.12).

After each section a summary will be given and some parts will be highlighted in bold which clearly structures this chapter. Therefore, an overall summary will not be given at the end.

4.1 *Notification*

The information in this analysis chapter is mainly based on the input from the Delphi-analysis. This input has been set up by questing ten experts in several rounds in which they can also respond on information from a previous round. The input is therefore refined during the several sessions. The first session consisted of six main questions and some sub questions. The input of each experts has been collected and scheduled at the appropriate question and answers which were related were combined or placed together. The basis of this chapter is therefore the first session. If there is no specification in the text from which session the information flows it is therefore from the first session. The following session has been built upon the first session. The first session has been restructured by adding comments of experts from the second and third session. The final text has been made by combining all the input and comments of three sessions. Some experts in the Delphi-analysis suggested websites or documents for more information, therefore this information has been written down in a textboxes in order to focus on first-hand input of experts. Furthermore, often quotes have been used because several experts often gave high-quality explanations or specific

opinions. Quotes also emphasize a personal perspective, which is highly important for mapping opinions and the search for consensus on open access to research data.

Also, interviews are conducted which may give a different or more in-depth view on open access to research data. Several interviews are discussed separately from the Delphi-analysis. If there would be too much overlap in the text, information from interviews was in some cases added to the Delphi-analysis section. In these cases a notification with a reference to the interview has been written down. There is also a clear distinction between experts and interviewees. Both are participants, but experts participated in the Delphi-analysis and interviewees in an interview. This clarifies the used method.

The theory from part one has been applied to the information from the Delphi-analysis and the interviews. This has been done by looking for differences and similarities between governmental strategies from the theory and the proposed governmental role by experts in the Delphi-analysis and interviewees. All in all, these information form together a perspective on the issue on open access to research data.

4.2 *Definition*

The Delphi-analysis started with the question: what is open access to research data? Experts elaborated by defining that data that have been generated or collected during the research process are made openly available for others (researchers and other) without any technical, legal or financial barriers and without requiring any prior registration. “Not only does research data need to be accessible (or ‘public’ data in general), but also the raw data that underlies the published research in articles or books”. The data that needs to be made accessible includes in any case the data underlying publications of research results in for example scientific journals and proceedings. One of the experts states that research results should really be considered part of the publication itself, for example “if this data is not made accessible then the publication should not even be accepted, for lack of verifiability”. This thought is considered interesting by another expert in the second session.

4.2.1 *Research data*

A question that arose during the Delphi-analysis and interviews is: what is research data? There is no consensus in society on what research data are (explicitly mentioned by two experts in the second session). According to experts, ideally research data are:

- raw data;
- methodology data;
- process data;
- processed data; and

- associated metadata.

A follow up question is: what is raw data? This has been questioned by one expert in the third session. Do we as society want to put “all data that comes from the telescopes in Chile in terabytes in the data centre or ASTRON in the open? Nobody will understand this kind of data.” It will be useful for others if it has been tidied up a little bit and put together with clarification on how something has been measured and what certain terms mean. See also more information in section 4.4.

“Concerning 'methodology' and 'process' data, these are not data in the strict sense but descriptive elements necessary to be able to access, understand and assess the data. According to one expert in the third session, “if process data is methodology then I see it to be separate but not new. One could argue that it is a necessity to be able to interpret the actual research data (output). So it is a requirement for both other categories and not a separate category in my view”. Methodology and process data are crucial elements that should always be provided together with the data (mentioned by two experts). According to one expert in the third session, processed data should be regarded as open only when working with a corresponding licencing scheme, but not all licensing schemes see data in this way. An example are the Creative Commons (CC) licences where the end result of the reuse of (open) data is also open, see more information about licences in section 4.9.6. **So it depends on what licensing scheme is used in which domain of research.** According to this expert, raw data and methodology data are less contested and should be seen as research data. Essentially, research data is dependent on the context and discipline specific (mentioned by three experts in the third session and by Mr. Hof, interview); “if its research in humanities or law it will constitute an entirely different set of results than in biomedical or economic research”. Additionally, processed data may not exist in some fields such as humanities.

Metadata provides information on key features of the data, “including the authors or contributors, the funders of the project in the context of which the data has been collected or generated, access restrictions and licences, persistent identifier and location”. For example, good metadata is crucial to allow for discoverability, proper attribution and acknowledgement.

One of the experts would like to change processed data into unprocessed. Or add the latter to the list of research data. In physics, like particle physics, for example large quantities of unprocessed data are being stored and in some cases only five or ten percent is actually used and processed into understandable data.

However, Mr. Spichtinger (interview) states that the definition of research data depends on the discipline. For example, researchers in the biology know themselves what is their research data. A

general definition for research data is therefore not possible, because it depends on the field of research.

Another expert in the third session disagrees on the aforementioned enumeration of ideally research data by stating that it is a definition of open science. Open science is a combination of open data, open processes and workflows (including open software) and open access to publications. Therefore, the scope of the given enumeration is too wide and cannot be applied to a definition of research data.

Other experts expand on the definition of research data and say it is about the facts that need to be interpreted (raw data, processed data, interpreted data, data coming out of a machine, a series of photos, of documents, questionnaires; mentioned by three experts in the third session) and data that have been collected or generated during the research process (through experiments, observations, measurements, simulations, calculations, surveys et cetera; mentioned by one expert in the second and two experts in the third session). So, it is about the building blocks that the researcher works with. “Data that has been collected in other contexts and is then used as input for research should not be included in an undifferentiated way under the header of research data”. For example, data that are collected by the national social security systems on citizens' contributions to the social security systems may very well build the basis for economic research, but it is not research data as such. All data that is a derivative of 'raw data' (for example: processed data) should also be considered as research data. Such processed data could very well be based on 'raw data' that have not been collected primarily for research purposes. “For example, if as part of a research project, social security data (which is not research data per se) is analysed and linked to data on migration that has been collected as part of a national census (which is not research data per se) to draw new conclusions, then the derived data would very well be 'research data', as it has been produced as part of the research process”.

On the following page, a textbox will be presented with two definitions of research data, based on the Australian Griffith University and the University of Minnesota (mentioned by one expert in the third session).

Research data definition

One of the experts in the third session proposes a definition of research data from the Australian Griffith University:

Research data are factual records, which may take the form of numbers, symbols, text, images or sounds, which used as primary sources for research, which are commonly accepted in the research community as necessary to validate research findings.

Another definition from the University of Minnesota, which focuses on several levels of data:

Research data are data in any format or medium that relate to or support research, scholarship or artistic activity. It can be classified as:

- raw or primary data: information recorded as notes, images, video footage, paper surveys, computer files, et cetera;
- processed data: analyses, descriptions, and conclusions prepared as reports or papers;
- published data: information distributed to people beyond those involved in data acquisition and administration.

These definitions make clear that the certification of digital archives is not only of interest to scientific archives of primary research data, but also for cultural heritage institutions such as libraries, museums and archives.

4.2.2 Open access

Experts also commented on the definition of open access. Basically, ‘as much as possible’ should be made freely accessible and available. The term ‘**open by default**’ should be applicable and means that research data should be openly accessible where it is appropriate while recognising the exceptions that need to be in place. By keeping in mind that some data are too sensitive for exposure, because of privacy reasons (sensitive information, embargoes) or security reasons.

One of the experts suggests the following definition of open access and this has been agreed upon by another expert in the third session (Open Definition, n.d.):

“A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike”

This definition has implicitly been supported by several experts; free usability and reusability are emphasized.

Importantly, not only the data from 'successful' experiments should be made accessible, but also the data from negative results (for example experiments that did not confirm a given hypothesis). This is confirmed by another expert in the second session, who states that failures are often hidden at this moment.

Instead of the aforementioned definition, one of the experts states that open access does not mean “free, without restrictions or payment”. For several reasons some form of (temporary) restrictions and payments could be desirable. A remark is appropriate in this context according to one of the experts: “data is already open as you can always pay people, there is a limitation to openness”. However, this statement is rejected by two experts in the third session, “open is free to use for any user: no payments for the user”. A policy advisor of the European Commission (Interview) says the same, the proposed it is not a common way, as it is illegal. Another expert in the third session states that if you need to pay to get access it is definitely not open (then similarly in the case of publications, subscription journals would also have to be considered ‘open’). Furthermore, in many cases it is not a question of paying for access, it is simply that there is no access provided to the data.

Additionally, data “which needs to be freely accessible is something as a forcing principle on someone else”, according to one expert in the second session. Enforcement to share is not appropriate, if people do not wish to share, they will find ways to circumvent. “A culture of sharing will result in sharing, not a contract”.

To know more about data that needs to be freely accessible, a non-exhaustive list has been made by experts in the first session:

- Raw data when possible, but definitely when processing is not done by generally accepted procedures;
- Data underlying peer-reviewed publications;
- Data which are hard or impossible to reproduce or for which generation was extremely expensive;
- Processed data, especially when this increases the usability compared to the raw data (for example a larger user community, use possible without extensive pre-processing, et cetera);
- All data that is necessary to correctly interpret and re-use the data and which enable verification;
- Metadata which will allow the easy discoverability of the data, provide information on the funder of the research from which the data originated, provide information on the ‘authors’ and other contributors, et cetera. Most important here are the data about the data: what are these data, what code is used, how can you use them and compare them with other data.

According to Mr. Hof (interview), metadata should entail information of what have been done with the data. So, have the data been checked, maintained and validated. This is in particular important for databases which have been created a couple of years ago or need to be stored for a long term (also mentioned by Mr. Hof, interview and Mr. Kolman, interview). Furthermore, research data can be presented in a blurred way (Hof, interview). In particular, this can be done in cases where a researcher does not want to present all his research data, because of privacy or safety reasons. The textbox at the end of this section presents more information about blurred research data.

Metadata

Metadata determines the usability of research data (Hof, interview). Therefore, metadata should entail all the information of what have been done with the research data. Is it complete, maintained, checked, validated or maybe blurred?

One of the experts in the second session made a remark on the aforementioned list by emphasizing the difficulty of the staggering amount of data which would need to be stored if all those data need to be freely accessible. This is repeated by one expert in the third session and also stated that networks and bandwidths (transportation of huge amounts of data) are a great concern.

An illustrative, non-exhaustive list of fields for open access to research data has been made by one of the experts in the first session (see below). However, experts made comments on that list in the second session and third session (five experts in total). One of the experts states that the given list of fields for open access is not a good selection. The list is too much focused on big data, patient data and public data, which is used for research purposes. The list neglects data from other sources such as small data that is generated in the lab in one small experiment, or data collected in interviews in social sciences projects. Additionally, “data extracted through painstaking working in some dark archives in the process of historical research are not included in the given list”. Experts in the second (one) and third session (two) states that the focus on fields is not desirable and a list of fields for open access to research data suggests that some fields are more important than others, which is not desirable. Another expert states that such a list is not necessary, it is meaningless because it is difficult to point out what field or discipline is relevant for open data. “It is all a matter of how it is funded (publicly or privately, and if public, then likely open), sensitivity (in regards to privacy issues) and more practical matters like demand for the data (from citizens/companies/other re-users)”.

Another experts in the third session adds that a focus on fields is useful in finding and analysing data, because then you have the comparable types of data together.

However, for the sake of completeness and as an indication of the focus of open access to research data the list is as follows with taking into account the aforementioned:

1. Health care, patient data (including that from clinical trials). Privacy issues make this problematic, but that is beside the point here;
2. Astronomy, data from astronomical observatories et cetera;
3. Particle physics, for example the Large Hadron Collider (LHC);
4. Engineering data, construction, maintenance data related to bridges, trains, airplanes, roads cars et cetera, for safety, fuel efficiency et cetera;
5. Environmental data, soil acidity, plant growth, air quality et cetera, these phenomena are not limited by region and require large amounts of data;
6. Climate and weather related data for the same reason as under issue five, but also for direct societal use, for example farmers, holidaymakers et cetera.

4.2.3 *Sensitive data*

As mentioned at the beginning of the section, open access needs to take into account sensitive data. Nevertheless, defining sensitive data is a difficult issue, according to two experts in the second session. Additionally, it shall become harder and harder to decide on what is sensitive data and the domain of 'sensitive' in this way only becomes larger as we are able to combine more and more data sets in nifty ways (mentioned by one expert in the second session). Perhaps the more important concern is how we should deal with different 'categories' of sensitive data. Two experts in the second session clarify different types of data. Data are sensitive if any of the following applies:

- Privacy concerns (personal data in the area of life/medical sciences, e.g. patient data, or data from social sciences projects involving interviews). It is about personal information or data that, when combined with other data sets, can be traced back to individuals (mentioned explicitly by two experts). However, one of the experts in the third session states that this is overstated. "It is much easier to get access to peoples information through other sources than a bio bank with anonymized data;
- Confidentiality issues, including for commercial reasons. (More specification of commercial data see below); or
- Security issues, in particular national security, where access to the data could be used for terrorist activities, for example. Also specific governmental data are involved in this security issue, for example strategic information could be sensitive. Additionally, atomic energy

information, chemical weapons information and lethal virus information are also sensitive security data.

A statement by one of the experts is that data which have been collected in longitudinal studies and data, which are privacy-sensitive or with IPR-issues should be accessible, with restrictions, but not openly accessible without any restrictions. However, one expert in the third session states that there are many types of longitudinal studies, for example if somebody observes the breeding behaviour of wolves in a certain region of Slovenia over a number years. The question rises why the data from this study should not be openly accessible? Additionally, opinions on what falls under 'sensitive' and what not are not clear and ever changing.

4.2.4 *Commercial data*

Commercial data, is also a very sensitive topic, as the provision of access to data from industrial research may have a negative impact on the competitiveness of the enterprise. Resistance from businesses and enterprises is mainly about publicly funded research. Businesses and enterprises are often involved in research what is partly governmental funded. If those companies need to provide their research data to public, then this could harm their competitive edge. In general, publicly funded research is any research that is funded in whole or in part by public funds, through public (tax) money (mentioned explicitly by two experts in the second session). Even if the contribution by the public purse is only small, it should still be considered publically funded. This can be national (through research councils and universities), or international (for example European grant programs like FP7 and Horizon 2020 or research communities like CERN). This also means that the basic requirements and principles concerning openness (open access to publications, open access to research data) should also be applicable to this kind of research that receives both public and private funding. However, the system should of course provide enough flexibility (opt-out options) to safeguard the legitimate interests of the beneficiaries and the other (private) funders, where full openness might have an adverse effect on these.

However, one of the experts in the second session states that if more than fifty percent of the project is funded by public money, it is publicly funded research. One expert in the third session mentions the same. In that case a contract cannot lock up all the data generated, only a sub-part, which is crucial to the private funders involved. The expert recognises that it will always be very difficult to establish the line between publically funded or not, it remains a grey zone.

As mentioned before, the provision of access to data from industrial research may have a negative impact on the competitiveness of the enterprise. This is independent of any questions related to the use that the company itself may make of the data (for example whether the results of clinical studies are used in the best interest of the patients, or perhaps ignored for commercial

reasons). Here it will be important to strike the right balance between the benefit for the research community or society as a whole and the cost to the company in terms of lost competitiveness. A research community in this context is a social unit of people that shares common values and common interests in the broad sense. They help each other and provide feedback on many aspects of the body of available data. However, one of the experts in the second session states that it is not easy to define a research community, because it is domain specific.

Later in this study the business perspective will be described more clearly by discussing the participation of the business in sharing their data.

NLBIF | Blurred research data

The Netherlands Biodiversity Information Facility (NLBIF) facilitates open access to biodiversity data from the Netherlands for everyone. “We ask researchers if they agree on publishing data in public domain”, according to Mr. Hof (interview). There is a choice of giving open access. Certain people chose to remove sensitive information or want to publish a blurred version of the research data. For example, information about protected or rare species will not always be given in detail. Data will be blurred by presenting the habitat as a big region instead of a particular location in a forest.

Blurring of research data can also be used in other types of research data, like in cases of sensitive or commercial data, according to Mr. Hof (interview). It is important that the process of obfuscating data will be described/mentioned in the **metadata**. For example, by stating: “all red list species have been removed for the file or displayed in a higher resolution” (Hof, interview). The usability of data depends on the purpose of research and the suitability of data. Over time, data should be checked, maintained and validated, this determines the **usability** of data. GBIS, the Global Biodiversity Information Facility, uses the term ‘fit for purpose’: what are the data and where can it be used for? This is highly important in open access to research data. Furthermore, data is still usable for other research, but this depends on the type of research. In certain researches, data should be detailed, otherwise it cannot be used. This is a consequence of blurring research data (Hof, interview).

Summary | The definition of open access to research data is ambiguous and difficult to establish. The term ‘open by default’ would be applicable and means that research data should be openly accessible where it is appropriate while recognising exceptions that need to be in place in cases of sensitive (privacy and security) and commercial data. Data that have been generated or collected

during the research process are openly available for others (researchers and other) without any technical, legal or financial barriers and without requiring any prior registration and can be reused and redistributed. Research data could entail raw data, methodology data, process data, processed data and associated metadata. Not only data from 'successful' experiments should be made accessible, but also the data from negative results. However, an overall and comprehensive definition cannot be found yet.

4.3 A solution to what?

There is not one single problem to which open access to research data can be the solution. According to one expert in the second and third session, nothing is the full solution to something. However, according to one of the experts in the second session, open access to research data should be seen as a *principle* and an *opportunity*. “Data are now underused and the IT-possibilities are now such that large data collections can be handled.” See also section 4.6 for more information on the opportunities of open data. Easier access to research data can contribute to solve several problems that the society is struggling with, such as:

- **Poor efficiency** (rather than reusing data that has already been collected in another context, experiments and observations are repeated and consequently funding is wasted). By opening access to research data duplication of research effort and data creation can be avoided and therefore costs. So, ease of access and re-use of data are one of the advantages of open access to research data. The more raw data is made accessible, the better, as this could then be analysed by other researchers to obtain new results, discover new connections and interdependencies et cetera. But of course these data need to be provided in a way that they can be reused reasonably easily by others, and it must be discoverable. So the raw data has to come with high-quality metadata, which will have to be created and made accessible (this has been emphasized by another expert in the second session).

However, the re-use of data moves the problem of open access forward as one does not know what will in the future be useful and how it needs to be stored or made available to meet such future needs and criteria as data curation and meta-data (mentioned by one of the experts in the second session). Additionally, one expert in the third session says that it does not “move the problem forward, since it is not practical to determine solutions to possible problems in advance. It is more logical to slowly or cautiously experiment; see which obstacles pop up and then act accordingly, instead of being paralyzed by anticipating too far ahead”. Unfortunately, this statement cannot further elaborated, since there is no additional explanation available.

Additionally, ease of access and reuse are entirely different than duplication, apart from the fact that duplication is indeed a repetition. One of the core tenets of modern science is the fact that it should be replicable, to demonstrate the veracity (mentioned in the second session by one expert). One of the experts in the third session makes a statement that this is an intentional repetition and not something out of ignorance. Again further explanation is missing;

- **Increasing number of irreproducible and unverifiable results** and in the worst case even **fraud** (especially in scientific areas that rely heavily on experiments, such as life sciences).
- The **difficulty to check, verify and test results** (mentioned explicitly by two experts). Open access to research data would facilitate the verifiability and strengthens trust in research. Validation and therefore evaluation of research will become much easier. This means that others are able to use, reuse and criticize the data. Questions rise such as: how did the research come to certain solutions? What is the data behind it? What is the methodology? Why is some data not included?;
- **Low level of transparency** and consequently **lack of trust** in the results of industrial research in some sectors. For example, pharmaceutical research data results may be not clear - without access to the data from clinical studies - whether new medicines are really more effective and have fewer side effects or whether they are only placed on the market for purely commercial interests. Open access to research data may combat scientific misconduct and foster the professionalization of researchers. Three experts in the first session mentioned this as a problem for which open access to research data is a solution. However, this will only help if the methodologies are without questions and the statistical analyses are beyond doubt (mentioned by one expert in the second session). Otherwise the data and the outcome will just move the debate to be about these elements.

Nevertheless, according to another expert in the first session: “open access will not necessarily reduce fraud or bad science as the people who use fraudulent data will usually be able to craft a data set which meets criteria for open access, as they are not restricted by actual outcomes.” However, one of the experts states in the second session that it is very hard to make up an artificial dataset that cannot be detected as such, so the ability to craft a data set which meets criteria for open access is rather difficult. “These datasets are too regular, outliers are too obviously added on artificially in an attempt to make the dataset look genuine. In addition, if open access aims to reduce fraud, there would be a need to ‘freeze’ the data as produced to create a ‘chain of custody’ and give the data the value of evidence. This is impossible unless it is not changed or repaired in any way.” And that is exactly what you want to do if you wish to create interlinking data sets. However, freezing

data is only an issue in the context of dynamic data, which will mostly be relevant in the context of big data, according to one of the experts in the second session. In the second session another remark has been made that data collection which intends to demonstrate fraud is by its nature entirely different from data collection intended to be linked and integrated in other data sets. “The former is more of a forensic type of approach, one would want to freeze the situation as left by the scientist, and not change anything”. If issues about the data arise at a later date, then the original data can be checked. Integrating data into another data set is about changing the data to match the wider set and this “disturbs the crime scene”;

- **Risk protections** and the **management** of those risks. It is about protection of reputation and the ability to meet the freedom of information requests. Unfortunately, experts did not provide more clarification and arguments.

Open access to research data is part of a more general ‘*open*’ movement (of which open source, open data and open spending are taking part in; mentioned by experts; Van Loenen, interview). “If there is a problem for which open access to research data is the solution, then the wealth of data available in governments (two experts explicitly state this), research institutes and (other) public bodies and organisations is the ‘problem.’” But these data only become interesting to take into account if and when they are used as input for research and publicly funded research (according to one of the experts in the second session). These research data were gathered with public means (tax money) or for ‘the general good’ (or in any case non-commercial ends) and many citizens feel that these data should therefore be freely accessible and many researchers, firms and people can benefit from the results. So it is partly an ethical problem of ‘**entitlement**’: people feel they have *the right* to the data and giving them this data solves the problem (this has been mentioned by two experts; Van Loenen, interview). This is however only applicable for publicly funded research and the research data that it produces.

“Another part of the problem is that many (societal and economic) issues are now tackled not only by public bodies, but also by *private actors* (mentioned by two experts). Granting access to (research) data would further help in enabling these actors to not only make money by building applications or a business on this data, but also solve societal problems. These problems might be issues that governments struggle with, but could be solved with many other actors contributing, based on freely accessible data.”

Summary | Open access to research data can be seen as an opportunity and needs to be approached as a principle. Easier access to research data can improve poor efficiency in research and foster re-use of data. Additionally, access may solve the increasing number of irreproducible and unverifiable results and probably to some extent even fraud. Furthermore open access to research data eases the check, verification and testing of results. A low level of transparency, the lack of trust and risk protection can also be increased by open access to research data. Accessibility is part of a 'open' movement and could be beneficial for tackling societal and economic issues in which public and private actors are involved. Open access to research data is also an ethical problem, because people feel they have the right to the data, it is about 'entitlement'.

4.4 *Necessity*

Every expert in the Delphi-analysis agreed that scientific (research) data needs to be accessible for the public, for each individual as it is about "democratisation of data". One expert states in the second session that everyone can start or own a business that benefits from research, as do people working in enterprises. According to one expert in the third session, some patient organisations are also very keen on having access to data. See for example, the Duchenne Parent Project.

"Research being sponsored by public means should be available to all, with the exception of sensitive data", is the justification (mentioned by three experts). Or: "in the end access is important to society which benefits from higher impact of research." One of the experts mentions: "the logical conclusion is that if the public is to trust research then having access to the data is important, but whether members of the public are key users of the data is another matter. The public will want to know that research is of sound quality and that it can be verified and that industry and policy decisions are based on full evidence." One of the statements is that the principle of 'open by default' is important. Open access to research data justifies governmental investments (Van Loenen, interview). In particular, periods of budget cuts people demand more accountability and reciprocity, open access to research data can be part of that demand for more accountability and justification.

Arguments for open access to research data outside of academia are that "**data is the new gold**", as European commissioner Digital Agenda Mrs. Kroes stated. Furthermore, the general sense of 'entitlement' can be seen as an argument for open access to research data (mentioned explicitly by one expert). "If there is no real sense of urgency in the citizenry to claim the data 'as their own', then citizens might not care enough to make an issue out of it."

Three experts state that accessibility of research data is foremost important to the research community, because these data can be used for follow-up research or re-used in other research.

Accessibility to all is desired by each expert. Accessibility to non-sensitive data to consumers (without specifying who that it), gives opportunities to individuals. Open access to data should be actively pursued, as sharing data can foster the advancement of excellent research in Europe. It may “bridge the innovation divide in Europe” (mentioned by one expert). Additionally, opening up the data to anyone may spur innovation and new insights. In order to increase innovation, the open access to research data must take into account the legitimate commercial, national security and privacy interest (stated by two experts).

Another remark from the second session is that it “is rather silly to expect that data is self-explanatory, there is nothing self-explanatory about it. It is always a **matter of context**. Any human endeavour is based on context. No technology, information or data can be perceived outside the scope of its linguistic context.” Furthermore, one experts from the second session states that “the nineteenth century liberal view that education will turn people in respectable and peace loving citizens is, unfortunately, not true. There is no alternative, but a blind faith in the ‘wisdom of the masses’ is rather naive. Apart from that, the sheer volume will make any intelligent use impossible, unless order is created in such a chaos. (!) In other words, **people will need to make decision about what and how it is shared**” and metadata are therefore important in certain situations.

Furthermore, scientific data will only have a value and be(come) information, to those who can understand it (two experts). According to one of the experts in the second session: “just as it is pointless to give an illiterate man a book, it is pointless to give the public unrestricted access”. However, two other experts disagree on this. One expert in the second session states that because of the indistinctness who can benefit out there, it is better to open it up to everyone. Additionally, one expert in the third session says that the phenomenon of valuing data and understanding it, could applies to many things, like financial balances or patents. The baseline needs to be: “the ones that can interpret it, or want to learn from it, need to be able to do so.”

One of the experts in the first session says that not all individuals understand data, however according to another expert in the first session their knowledge cannot be underestimated. “Citizens act in two ways – creating data by crowd sourcing or annotating and consuming data”. A general notion from one of the experts is that most so-called ‘ordinary citizens’ will never in their life need or even want to have access to any kind of research data; it would not even occur to them. “There will be a few individuals who both want and need access to scientific data”:

1. be it because they are carrying out research on a specific topic outside of any institutional framework (‘hobby scientists’ who are actually doing serious research); or
2. be it because they are engaged in citizen initiatives about specific topics (for example environmental issues where political decisions rely on scientific studies);

3. because they belong to an industry: they are in fact a collection of individuals looking for information to boost their innovation.

In the first and second cases “getting access is important for the individuals concerned, but not only for them – if the ultimate outcome is new scientific knowledge, then this can potentially be important for society as a whole”. However, one of the experts in the third session states that it is not an argument not to open up the data in cases where ordinary citizens will not need or want to have access to research data.

(!) *One of the experts states that it is easier and more cost effective to open-up to all than implement and maintain a very complex federated authentication/authorization mechanism to give only access to the scientific community as a whole, groups or specific individuals.* Therefore, opportunities as eduGAIN exist, see for more information the textbox below. Another expert in the second session agrees on the aforementioned statement, but questions how we can decide who can have access and who cannot? Specifically, SME’s often innovate and benefit a lot from open access to research data. Mr. Hof (interview) mentioned that it is difficult and technically impossible to exclude individuals or (commercial) organisations. Because it is not feasible to establish two domains (public and commercial use). “If you do not want to give access to your data, it will be better to do not publish it online”, according to Mr. Hof (interview).

eduGain | Access to specific groups

“eduGain interconnects identity federations around the world, simplifying access to content, services and resources for the global research and education community.” An identity federation is a group of institutions and organisations that sign up to an agreed set of policies for exchanging information about users and resources to enable access to and use of the resources.” eduGain enables “secure exchange of information related to identity, authentication and authorisation between participating federations by coordinating elements of the federations’ technical infrastructure and providing a policy framework that controls this information exchange.”

Information in this textbox has been derived from the website of GÉANT, a project of which eduGAIN is part of (GÉANT, 2014).

4.4.1 What if not?

What if there is no open access to research data? According to the experts open access is a movement which is here to stay and “we should strive for it”. There is already a lot of data available, which has helped our society in many ways. Think of the human genome data (cancer research, biology) and data in the geosciences (climate research, meteorology). Data are in principle (re)useful

and could be open and made available if possible. The great difficulty here is who decides what is 'useful'? Probably not all research data will be open in the end. It will go step by (little) step, taking into account technological and financial restraints. One of the experts welcomes a more modest approach of focussing on data underlying publications. "It will enable us to remain within the existing academic mind set (for example publications) and create better and more relevant output without the ambition to change a mind-set and create a new system." According to the expert, better and more relevant output "sounds nice as it is all outside the bounds of current practice, but has so many variables that it is very hard to make it work." Additionally, important issues concerning this 'new world' still are not addressed but should be, more prominently, in parliament and society at large (for example questions about ownership of data, or the value we place in data, et cetera; see also further sections in this chapter).

No open access will not make such a big difference for the **general public**, especially the non-specialist type (mentioned by two experts in the third session). "The requirements to really enable users, especially the non-specialists type, are tremendous." Besides, they will miss perhaps the possibilities that third parties take in developing software translating data in something useful. An example of useful public data is a new-developed application for electronic devices that show all public toilets in Amsterdam (mentioned by one expert in the third session). For researchers (and all parties that would like to do research) it would mean less possibilities to have a broader dataset and to use data from other countries. Reanalysing data of an article to see if you get the same results with your methods will not be easy without available research data. One of the experts states that his concern is to what extent "it will create a new divide between those who own processing power and those who do not."

Summary | Research data need to be accessible for the public, for each individual. It is about "democratisation of data", especially research sponsored by public means should be available to all, with the exception of sensitive data, according to several experts. However, research data are probably foremost important for the research community. Ordinary people perhaps do not understand and are not interested in the data, probably only few want to have and need access. Additionally, data are not self-explanatory and are a matter of context. A suggestion is that opening-up the data to everyone is easier and more cost effective than a complex system which gives access to specific actors, but a modest approach is also suggested. Open access to research data is a movement which is here to stay, however questions rise such as: What data are useful to open-up? and, how do we decide what is useful?

4.5 *Ownership*

The discussion who is the owner of data is ambiguous, according to the experts and interviewees (Grosfeld; Hof; Kolman; Spichtinger; Van den Biesen). It is not a very productive discussion and ownership over the data is not desirable nor any exclusive rights, mentioned by one expert in the second session. According to Van Loenen (interview), ownership is an important issues to think about, but not the starting point of open access to research data. More important is the question: open access to research data or not? Including agreement on (basic) definitions. More harmonisation of ownership would be desirable including an offensive approach. The awareness of open access should be increased, however this takes a lot of time.

Two experts in the second session state that data as such have no owner, although contractual obligations can determine who controls it. Data merely have a controller. Ownership rights do exist in intellectual property, such as database rights, patents, copyrights, but these property rights are granted by law and there is no comparable law which grants ownership to data. The use of the word ownership will not make it a legal fact. "Privacy has no bearing on ownership, similarly neither do rights concerning the integrity of for example the body to tissue or organ transplants. This is not a question which is up for a vote unless it is in parliament." Additionally, one expert says that it is not possible to outright define the owner of data.

Accessibility to data needs to be given by the 'owners' of the data, but this raises a lot of debate. The general opinion of the experts is that access to research data should be given by the institute and the researcher. "A 'research community' should give open access (providing necessary contextual information), research funders and stakeholders that benefit from open access should contribute to maintaining open access (providing incentives and resources)."

Mr. Kolman (interview) also notices the ambiguity of data ownership. He states that universities, funders or individuals claim 'the right' of ownership. However, clarity does not exist yet.

Researchers can be seen as the owners, but in reality it is usually the institute that the researcher works for that owns the data, together with the funder of a study (based on the expert panel). To be more precise, entities who generate/use data (government, pharmaceutical companies et cetera) and scientist/researchers who generate/use data do need to have access to research data. Additionally, one of the experts states that researchers do primarily need to give access to the data, because they generate and collect the data. Furthermore, the employers of the researchers and the funders of the research projects also need to give access to the data. "They should encourage and support researcher in making their data easily accessible, be it through practical and technical support or through financial support."

However, according to one of the experts in the first session, the contractor owns the data (even legally) in the case of industrial research or contract research (for example, research carried out for public authorities) and the one who should provide access. Nevertheless, this depends on the contract, according to one of the experts in the third session.

One of the experts in the second session states that ownership depends on the perspective you take; a legal or a moral perspective. A **legal perspective** will be quite clear in individual cases, but will differ between jurisdictions. Clearly only the legal owner can provide access to the data, unless the owner has delegated these rights to a different entity or individual. According to this expert, who is not a legal expert, research results in some European countries generated by academics and researchers during their research job at an university are owned by the institution, not by the academics themselves. In those cases it will be important that the institution grants academics the right to disseminate the results (including research data) in an appropriate way; clearly defining what can be disseminated and how. This expert mentions that this maybe is perhaps already happening in many cases.

A **moral perspective** supposes that the owner of the data should be the person or team that has collected or generated the data (mentioned by three experts in the second session and Mr. Kolman, interview). The person or team who created the data has some kind of ‘emotional bond’ with the data. “They are the ones who will best understand what the data actually means, how to interpret it, how it can be reused et cetera. So they should also take on the responsibility of making it available to other in an appropriate format (including for example proper metadata). However, there is no legal basis for this. “The emotional link between a creator and the outcome of his work are of course very real, there is just not a right of ownership”. Another expert in the third session adds that one could argue that it is not the collector who is the owner, but the object (person) that the data is collected from or about. For example, the citizen in case of data pertaining to citizens, like addresses or personal public service number (the Dutch BSN number).

However, the discussion of the ownership shifts to a situation of co-ownership (according to one of the experts in the first session). Ownership may very well be shared by different organisations. The organisation who is registering the data, the researcher and the actor whose information is being registered, the research subject, the main processor of the data and the re-user of the open data. In each case the ownership has to be defined on a more or less ‘ad hoc basis’ as long as there is no general way of deciding on ownership. However, the expert who is proposing this, is not sure if such a general way of deciding ownership would be feasible or desirable.

One of the experts in the first session says that in fact the institution that employs the researcher is the owner of the data. In the academic world the owner of data is the university (mentioned by one

expert in the second session). But this is different if contracts have been made with private companies. The owner in this context can give permission for open access, considering privacy, contractual and copyright restrictions.

An example of an institution that is the owner of data, is the researcher funder 'Netherlands Organisation for Scientific Research' (NWO). In cases where data are collected in a project that is funded by NWO, co-ownership will be claimed. In this case of contract research the data are initially the property of the contractor that has generated or collected it, but in many cases it will be part of the contract that this data are then handed over to the entity that has awarded the contract, as one of the deliverables of the contract (mentioned by one expert in the second session). Then it of course becomes the property of that entity that has ultimately paid for it. In this case the situation is different from what was written before – "the data have been generated or collected as part of a 'product' that was to be delivered (for example a study on a certain topic), and once the 'product' including the underlying data has been handed over, the contractor has fulfilled his job and it is up to the entity that has commissioned the work to decide what to do with the data". On the other hand, if the handing over of the data is not part of the contract and there are no legal constraints (for example confidentiality) then the contractor remains the owner of the data and should provide access to it.

Another way to appoint an owner of data could be by thinking in terms of 'main processor' of the data, since this may be more recognizable in practice and perhaps a bit more clear to define than ownership (according to one expert in the second session). Here the main processor role may also be divided over more than one party, but this 'simply' reflects the complexity of the subject matter and should be acknowledged as such.

"In this kind of set-up, the party being 'registered' (for example the research subjects) would ideally be given the brunt of rights over the management (and quality control) of the data and would also be mainly responsible for authorizing access and should be given tools to ensure the quality of the data. While other processor roles (researchers, other re-users) would more narrowly be subjected to rules or duties related to proper use of the data while taking account of the 'framework' set by the main registered party."

However, according to one of the experts in the second session, the emphasis should not be on authorizing access all too much (but the registered party would be responsible for that, if anyone) but rather on quality control and management.

In general, ownership should be defined in the Data Management Plan (see section 4.9.2; Hof, interview). This results in a clear definition of the owner and clarifies the legal issue of ownership.

Ownership has been discussed and defined in the *RECODE project*, mentioned by one of the experts in the second session. A copy of this perspective and scope can be found in appendix G (Finn, Wadhwa, Taylor, Sveinsdottir, Noorman, Wyatt & Sondervan, 2014). In the report a distinction has been made between ownership of and authority for giving access to data (latter will be discussed in the following subsection). The intellectual property rights are also discussed and entails the protection of creative works by individuals that are the result of innovation, skill and specialist effort. “The governing of intellectual property rights in relation to open access to research data references both moral rights and exploitation rights for the researchers who created collected or curated the data”. “In relation to open data, database rights prevent third parties from publishing, distributing and copying research data.” This study will not go in-depth into intellectual rights, as it falls outside the scope of investigation.

Summary | Defining ownership of data is ambiguous and depends on the perspective taken. Ownership in legal and moral perspectives differ in thinking in terms of ‘main processor’. The general opinion is that access to research data should be given by the institute and the researcher. However, co-ownership and contracts often define the authority for giving access to research data. Ownership should be part of the data management plan (DMP).

4.6 *Opportunities*

Open access to research data is not a solution to certain problem as mentioned before, but open access to research data does have opportunities for society. According to one of the experts, benefit depend on the goal of open access to research data. For example, companies may benefit from data sharing, but not at the expense of their competitive edge.

Open access will (summary of all participating experts and interviewees):

- **Increase** the **quality** and **efficiency** of scientific research (mentioned by seven out of ten experts). “Open access has the potential to provide all stakeholders with evidence of the high standards of quality and integrity the scientific system has imposed on itself”. This is highly applicable, because the publication pressure in the scientific sector focuses on the quality of the paper, not the data. However, data should not be based on faulty or fraudulent data. The quality may increase because of new opportunities and improvement of the efficiency and effectiveness of research investments. Additionally, it may affect positively the way research is designed and carried out which makes research more grounded. “Experiments will be better planned and documented, when data can be scrutinized and checked by others”. The

efficiency can be improved, because researchers can built upon existing research more easily;

- Facilitate **research integrity**, particularly the validation and therefore evaluation of research (results). It demonstrate the results of research and how conclusions are drawn. Validation of the published results is possible because of access to the underlying data. Furthermore, other people are able to use, re-use and criticize the data and open access allows to verify data, reproduce and test the research conclusions. Because of this, new insights in research processes can be obtained due to existing “knowledge-silo problems”. Research transparency is related to research integrity (mentioned by five experts); an increase of transparency of the practice of scientific research and its results in the form of publications and the underlying research data affect the research integrity positively;
- Create **trust** in and **legitimacy** of research organisations and research findings and if presented in the right way it can support public understanding (also mentioned by Van Loenen, interview). Because of this legitimacy public engagement and societal accountability will increase. This is, according to one of the experts in the second session, again a ‘data speaks for itself’ argument. This statement is underpinned by saying that “if it were true there would not be politicians and especially no populists. However, as there have been a feature of European politics since ancient Greece, this seems rather untenable.” Additionally, data can always be reinterpreted, methods can be questioned or the suggestion of conspiracies, lies and other malfeasance, not matter what you share. According to one of the experts, trust is therefore between people and not between data and people. “If you trust your local Member of Parliament over scientists, it will not matter what is shared”. Another expert in the second session also states that many people speak about trust and suggest that it is data that people trust. However, the expert considers that this is nonsense, because people trust people, so the origin of the data determines the trust it receives. “The quality of the originator or creator is the deciding element. The data itself can only help to build such trust if the reader is able to judge the competency of the originator/creator. Data as such will not help anyone without the pre-existing knowledge about the field, requirements, state of the art et cetera. In other words, someone who is aware of the context will be able to value data, people who are not, will not suddenly become aware of such a context, neither will it help them to trust this (unknown) context.”
- Boost **re-use** and **data sharing** in order to stimulate creativity and the discovery of new fields of research (mentioned by five out of ten experts). The open access gives access to more data and combining datasets creates larger datasets that are in principle more statistically robust (mentioned by three experts). Another advantage of sharing and combining different

datasets is the possibility to answer other or more (research) questions than would be possible to answer with the separate dataset only. Consequently, more research may take place across research fields. Furthermore, new ways of assessing research come up, because the influence on society can be noticed by analysing how the data has been re-used;

- Give businesses and industries (SME's) access to **support innovation** (mentioned by seven experts) in order to uplift the economy and to secure the welfare of a country). Barriers (as financial and registration) to the data make it harder for SME's to innovate. "Serendipity does not happen when everything is locked up." Open access can contribute to the economic growth (also mentioned by Van Loenen, interview). Additionally, an increased recognition of the value of the work that goes into data curation and data management may also lead to new (recognized) jobs in that area (according to one expert). Data citation and data publishing could also be upcoming job disciplines. "Currently many of these tasks are carried out by researchers 'on the side', not always in the most efficient way. Training specialised data experts (and creating the demand for it in the first place) would increase efficiency and open up new career pathways". However, one expert in the second session states that most problems may not be solved by merely finding the other end of a spectrum. "If you are in the desert, you will die for lack of water, but if you are at sea, you will die because of the sheer amount, even if on a boat as it is undrinkable";
- **Widen participation** of the latest European Union member states. Without open access data cannot be found, no new opportunities and partnerships can arise. But this opportunity has been defined a bit far-fetched in the second session. According to one of the experts, open access to research data in this broad context could then also be related to development aid and capacity building in the industry. So these issues will be a bit far-fetched advantage of open access to research data.

Summary | Open access to research data may increase the quality and efficiency of scientific research and facilitate research integrity. Moreover, it may create trust in and legitimacy of research organisations and research findings. Accessibility may boost the re-use and sharing of data and probably to some extent widen participation of the latest European Union member states. The main opportunity is to support innovation by giving businesses and industries access.

4.7 *Difficulties*

Open access to research data gives opportunities, but there is always a (potential) down-side. "If that is ignored, the person who moves last reaps the biggest benefits". However, leeching or free-riding

does not form the basis of a vibrant culture of data sharing. Difficulties in open access to research data are (summary of several experts):

- That the **researcher that created the data usually wants to hold on to them**, for different reasons like (the situation may be different in each discipline and it can change rather fast):
 - It is my data that I worked very hard to collect, and no one else has the right to it;
 - If someone else analyzes my data, they may come up with a different answer disproving my perspective;
 - I have not finished analyzing my data, and I will make it available once my analysis is complete;
 - I cannot trust the data produced somewhere else;
 - I have a life (family, mortgage) which depends on being able to carry on with my research and not be overtaken by bigger labs, or start all over each time;
 - I want to write a few more papers on basis of that data.
- “**Lack of awareness for the benefit that open research data could bring to the science system**”;
- “**The fear of everyone being able to see what goes on in one’s organisation**. This may concern both public and private organisations”. As said before, private organisations (companies) may fear that their competitive edge diminishes as soon as they provide more data about their own organisation or data they own in general. But research institutes and public bodies may also experience fear or hesitation, because publicizing data and transparency exposes also the faults within an organisation (both fear and shame may be reasons to resist all-out publication of data). Additionally, “some researchers fear that making data available for others will give competitors an advantage in the race for funding and career advancement”;
- Issues related to **privacy, confidentiality or security** for some kinds of data could be constraining, see therefore section 4.2.3;
- **Legal obstacles** and uncertainties, especially in data mining¹;
- **Costs** of publicising data, especially the costs for uploading (transforming) data in a database;
- The **infrastructure** of the data; the *technical* aspects (mentioned by three experts). Depositing and opening data means a lot of storage to keep the data safe. The access and maintenance of data is also part of technical barriers. Additionally, the technical issue of publicizing data comes up when some systems may require more *investments* than others to be able to give the access desired (especially old/’legacy’ information technology systems) or

¹ Experts did not elaborate the issue of legal obstacles. Therefore, future research is recommended.

some institutions may have lack of support to opening up and storing the data. Specialized *personnel* may also be necessary in order to establish robust data infrastructures which optimise the way of dealing with data. Additionally, *software* that might need to be used to interpret and understand the data needs to be available;

- Related to the aforementioned aspect: **lack of training**, “especially young researchers fear that making data available for others will give competitors an advantage in the race for funding and career advancement”;
- The **lack of clarity about standards, licensing and protocols** to be used (mentioned explicitly by three experts). “This is a mixture of at least a regulatory and a technical problem; lacking clarity on these issues, it is not clear to organisations whether and how they should publicize the data (which data sets, what format, how frequently, under what – if any – licensing regimes, etc.)”. Creative Commons (CC) licences can help, but according to one of the experts that is probably not sufficient. See more information about licensing in the following section. Harmonisation of contradictions in open access between funders and the definitions in use are necessary. Additionally, the lack of incentives, recognition and harmonisation may constrain the open access to research data. Without proper recognition in the context of evaluation and appraisal many researchers will prefer to spend their time on other more rewarding activities;
- Data cannot be easily verified, tested or deeply understood without **context** and provenance information with the data (mentioned by two experts). Questions rise as: what to do with metadata? How to organise that? What is this data set? Who worked on it? How to cite it?;
- New **(research) communities** will be created, but **implementation** is lacking. An advantage is that these communities may facilitate the exchange of information in that community. However, practical implementation of or experiences with these communities are unclear, because new communities formed outside the bounds of existing groups, “as the data will be formed in a manner consistent with a certain ‘tribe’, makes it unlikely that it is (easily) usable for others.”

Summary | Accessibility to research data may also have negative effects. First of all, the researcher himself could refuse to give access and there is a certain lack of awareness of the benefits of open data. People who are aware, are afraid that everyone will be able to see what goes on in the organisation. Data related to privacy, confidentiality and security issues may also hinder accessibility, just as legal obstacles and costs of publicising data. A great barrier is the contemporary infrastructure data. Access needs to be given to a large amount of data which needs to be stored and

maintained. Furthermore, software for interpreting data needs to be available. Personnel is not always properly trained in dealing with data and there is a lack of clarity about standards, licensing and protocols. Additionally, data cannot be easily verified, tested or deeply understood without context.

4.8 *Intervention*

Because of the difficulties in (the process of) open access to research data certain interventions are suggested by experts in the Delphi-analysis and by several interviewees such as Mr. Hof and Mr. Spichtinger.

Delphi experts suggest that there is a need for intervention at several different levels and in parallel. “Simply inventing new mandates and regulations without also creating a support framework will not work at all”. Intervention through governmental actions could stimulate the move towards more accessibility to research data and may speed up the adaption to more open approach to research data. But this ‘push’ should mostly be targeted to realise accessibility of research data to the public and not directly to realise innovation or steer the way data are being applied by the re-users and scientific community, the drive needs to be about practical aims instead of principles (mentioned by three experts). The debate is not about how and why, but “openness is good for the sake of openness”. “Only when this community stalls or does not deliver could the government try to stimulate innovation” and when things go smoothly, it is not necessary to interfere, according to one of the experts in the second session and Mr. Kolman (interview). This is more legitimate than intervention by default in the sense that the government does not fully get involved in the private/commercial domain. Therefore, governmental facilitation is preferred which creates a more level playing field. “Creating momentum to form a community around an (existing) dataset, be it from entities and/or researcher, will require time and money, and as the latter buys the former, subsidies might be a way to create a tipping point”. Open access seems to be required in cases, which are funded by public money, so intervention by the government in this sense seems to be appropriate. Leading initiatives should help to set standards and would take care of interoperability. A government should stay informed, because unexpected regulatory factors may play a negative role in the success of open access initiatives. All in all, **the government should take up a role as a watchdog and show responsiveness instead of taking a role as director of open access to research data**. The governmental role should not be quality control, access and additions, according to one of the experts in the third session.

Mr. Van Loenen (interview) describes possibilities of the governmental by explaining four roles: (1) regulation, (2) financial measures, (3) purchasing power, (4) fostering networking, data sharing and awareness. The first and second roles are often mentioned by several stakeholders, but

in this case regulation in the sense of legislation is not appropriate in the case of open access to research data. However, the government should have a discretionary approach in open data and stimulate the share of experiences, according to Van Loenen (interview). Putting stakeholders together and to come to solutions initiated by the stakeholders themselves, that is what a government should do (Van Loenen, interview). The formulation of definitions and joint solutions should be the same. Stakeholder should understand each other in order to get convinced and cooperative. Therefore, the last mentioned role would be very important, according to Mr. Van Loenen (interview). However, not all the experts do fully agree on these interventions.

In the third session one of the experts of the Delphi-analysis explained very clearly two ways in which the government can and should intervene in an appropriate way:

1. In its role as **funder**, at national or regional level through its own agencies or by providing funds to (semi-)autonomous national funding councils, but also at European level by influencing the rules applicable to European funding schemes such as Horizon 2020. Here a mixture of 'carrot' (additional funds) and 'stick' (non-eligibility of costs if certain requirements are not fulfilled) can be used. However, it is important here not to forget the 'carrots' by focussing too much on the 'sticks'!
2. In its role as **legislator**, by providing a supportive legal framework – again the 'carrot' – or by imposing certain legal requirements on relevant stakeholders – the 'stick'. This could for example be a requirement that all research data produced within public research institutions have to be made open in an appropriate way, irrespective of the funder of the research, or a requirement that every publically funded research institution has to have a certified data repository that fulfils certain interoperability standards and is linked to a national data centre, et cetera, or that every recognized programme for a Research Master or a PhD has to include a module on research data management.

The main purpose of intervention is to set up facilities for a researcher to manage his data in a proper way. Therefore a certain framework of funding, tools and storage is necessary. To sum up:

- Funders need to demand data management plans from the researchers they fund;
- Funders need to demand long term storage of (a selection) of the data in trusted digital repositories;
- Funders need to supply funding for data management activities by the researcher.

Several aspect play a role in the debate on intervention of open access to research data. In general, rules for open access to research data or even filters in databases cannot be avoided (Hof, interview).

Nowadays even robots can damage data systems or hackers can remove billions of data, this is a burdensome for the system, according to Hof (interview).

Intervention (without specifying what kind of intervention) can also be done by universities (Hof, interview). However, the transition to open access to research data is difficult for universities; in general they do not know what to do (Hof, interview). “Ten years ago I talked about the same topic and there are not many differences now” (Hof, interview).

Mr. Hof (interview) and Mr. Van Loenen (interview) mentioned that broad European rules could be necessary. If many stakeholders do not give open access and each stakeholder determines own open access policies, the goal of efficiency cannot be realised. Reciprocity can be established via the European Union, because all the member states will follow the same rules or guidelines. Mr. Spichtinger (interview) mentioned that stakeholders have to work together. In particular, the supranational government with national and sub national governments in incentivising open access to research data. The dialogue between stakeholders is essential for establishing open access to research data. Additionally, how to deal with intellectual property rights and its implementation. The EC pilot in Horizon 2020 on open access to research data gives a clear signal to stakeholders. “That is what we can do now at European level” (Spichtinger, interview).

However, Mr. Kolman (interview) hesitated whether the government should take up an intervention role. **If the process of open access to research data can be arranged via stakeholders, then the government should probably not intervene.** Mr. Kolman supposes that proper datasets per discipline will get the norm in a few years and that therefore governmental intervention is maybe not necessary. Although, funding for data depositing, data maintenance and discoverability could be a governmental role. Data which need to be stored for at least fifty years would be costly, a government can therefore finance these data infrastructures, according to Mr. Kolman (interview).

In general, experts and interviewees gave suggestions for interventions at several levels, which can be specified at governmental, institutional and funders level.

4.8.1 *Governmental level*

Interventions at governmental level can be (summary of experts and interviewees):

- **Establishing a suitable framework for open access to research data and remove legal obstacles.** In particular, safeguarding the framework is essential a governmental intervention. More information about a regulatory framework see section 4.9. For example, governments should ensure adherence to appropriate licensing schemes, ensure that sensitive data are either not completely open or surrounded by the appropriate safeguards

and obstacles to protect privacy (for example access control). However, one of the experts explicitly states that access control is not a governmental role. Additionally, government should provide a reliable budget stream for the framework (mentioned by one of the experts in the third session). “Only then can you develop a sustainable network, only then can you get the trust of researchers.”;

- **Legislative measures.** There is a necessity to tackle the legal obstacles and uncertainties (as for example to data mining) through legislative measures. “Legislative intervention would also be needed to tackle those case where legitimate requests for access to research data are denied (for example by pharmaceutical companies not giving access to the data from clinical studies).” Other legislative measures are the reform of copyright and possibly a digital archive for the content of the institutions;
- **Financial support,** because the costs for data depositing and sharing can be high (also mentioned by Mr. Kolman, interview). Starting subsidies to help fund the investment in infrastructure is vital as it is something which is not feasible on a large scale by the institutes and as it is not a commercial offering, there is no market to turn to. However, one of the experts states that financial support can also be done through funding agencies as institutes and universities;
- **Policies to stimulate open access to data by giving incentives.** For example, incentives that promote the possibilities of innovation through open research data in business, enterprises and research organisations (also mentioned by Mr. Hof, interview and Mr. Spichtinger, interview). The focus should be on the creation of awareness and incentives for researchers as the data management plans and credits. However, one of the experts in the third session does have less faith in these ‘incentive policies’, since rewarding academics is not done by the government. Science is international, local reward schemes are therefore pointless;
- **Create an environment where people (researchers, librarians and repository/data manager) can be trained properly.** “We tend to think that is easy to set up a DMP or offer just a storage facility.” But it is more difficult to establish a DMP (researchers struggle with establishing a DMP), because criteria are ambiguous and the purpose of a DMP is not common ground yet (Hof, interview). Training would stimulate open access to research data and a correct use of a DMP by following certain steps during a research. See more information about these steps in section 4.9.2. In addition, Mr. Kolman (interview) mentioned that researchers hesitate to share their research data;
- **Bringing together different researchers and disciplines.** Important is that all stakeholders know what each of them is doing or can be asked to do, they may join forces to tackle problems or act together in for example stimulating open access;

- Assistance in providing access to data for a more general population by **informing** about which infrastructure could be used and which IT-solution works best;
- **A more (heavy) regulatory role in disciplines or sectors where the supply of open (research) data is lacking** (the geo sector is mentioned as an example by one of the experts). This is in situations of mismatch between the demand and the supply. If there is lacking demand, no state intervention is necessary because enforcement by regulation will not convince people of the benefits of using and requesting open data. In addition, this whole process of supply and demand of open data is more articulate in the geo-sector than in other sectors (according to one of the experts in the second session). However, the expert is not sure if it was processed correctly; and
- **Ensure that stakeholders cannot create a monopoly on open access to research data.** “A dominant ‘information’ position by a group of individuals should not be abused and access should not be refused on discriminatory grounds. Such a framework should not be technology driven, as it would become obsolete the day it is conceived, but be based on the normal principles of law, as set out in the laws mentioned above and other related laws (mentioned by one expert). In the end it is about human interactions and relations.” Another expert states: “under no circumstances should a small group be able to hold out or leverage a data set which a government would require access for safety or reasons of public morality. Of course data sets may offer a commercial advantage so anti-competition law (and obviously state aid restrictions) also plays a role.”

4.8.2 *Institutional level:*

Intervention at the institutional level can be summarised as follows (based on the experts):

- Open research data should be taken **into account for evaluation, promotion and getting a doctoral degree**;
- **Anonymising data** in order to make it possible to open up the access to some of the data. This should be done at the source by the research team;
- **(Meta)data quality control.** For example control of data in certain discipline, used formats and the legislation/rights. However, one expert in the second session questions who should control these (meta)data? Clarity on this issue is missing. However, another expert states that the government is responsible for the quality control, but in practice a specified agency should be responsible for this. Still, in time, controls of field and used formats could be ‘built in’ in the technology itself, which would mean a large role is also reserved for developers of relevant Information Technology (IT) systems. Another expert states that in practice part of metadata quality control is done by data repositories. Depositors of data are required to

follow a prescribed metadata scheme and use a limited number of preferred formats for the data they want to deposit;

- The institution will **provide workflows and processes** to include content in either their own repository or a national / subject-specific one. Unfortunately, more in-depth information became not available in the Delphi-analysis and interviews;
- The institution needs to **invest in support** for researchers, technically as well as in staff (advocacy, expert advice on legal issues, technical information, metadata, DMP et cetera);
- The researcher needs to be able to **trust** the framework by participating in an environment of sharing data. This can be achieved by creating **trust** in and **legitimacy** of research organisations and research findings. It needs to be presented in the right way in order to gain support. See more information in section 4.6; and
- **Provision of conversion tools** to open formats and tools to choose the right licences.

4.8.3 *Funders level*

In general, funders should according to multiple experts:

- **Demand** open access as a requirement for funding;
- Need to **set specific policy standards** (nationally as well internationally) using **soft law** measures and in some cases **mandates**. Mandates will be useful, but only if they are monitored properly and if sanctions are actually enforced. “Many funders have policies regarding open access to publications, but increasingly also regarding research data, but most of the time their knowledge as to the level of compliance is very limited due to a lack of effective monitoring, and very rarely are any sanctions imposed. This is well known and the risk of actually getting penalised for non-compliance is extremely low, making such non-enforced mandates a very ineffective tool”; and
- Play an important role in **coordinating** the stakeholders. Stakeholders have to work together towards a common goal (mentioned by two experts). A common drive needs to be set up by people who have a vested interest in creating and maintaining the dataset. Furthermore, funders could play a coordinating role in the data underlying publications, just like journals and publishers. For example: “data policies as Public Library of Science (PloS) which requires all underlying data to be deposited and the location to be communicated, can be a very effective way of getting researchers to comply, as non-compliance means non-publishing”. PloS is a “non-profit publisher and advocacy organisation founded to accelerate progress in science and medicine by leading a transformation in research communication” (PloS, n.d.1).

4.8.4 Publishers

The role of publishers has not been mentioned by the experts in the Delphi-analysis, therefore this section about publishers is very short. However, Mr. Van den Biesen (interview) mentioned that there could be a role for publishers in the open access to research data process. Mr. Kolman (interview), describes that publishers can help researchers to deal with research data, in particular by helping to describe data (metadata). For example, Elsevier can help to digitalise data and help researchers to record data. More information about possible help with data storage by publishers can be found in section 4.9.4. Mr. Van Loenen (interview) mentions that publishers are highly important in open access to research data, because publishers facilitate the peer review function. “They are probably more prepared for open access to research data than universities, because they need to transform their complete infrastructure systems” (Van Loenen, interview). Publishers are very important in the open data debate, but are often not taken into account or people do not recognise the value of publishers.

Summary | Some form of intervention is desirable according to the experts and several interviewees, however this depends on individual point of views. The main purpose of intervention in general is to set up facilitation for a researcher to manage his data in a proper way. Therefore a certain framework of funding, tools and storage is necessary. Several examples of interventions are proposed at governmental, institutional and funders level. The government should take up a role as a watchdog and responsiveness instead of a role as director of open access to research data. However, its role as funder is often suggested and even as legislator by providing a supportive legal framework. Bringing together disciplines and researchers and providing training are also suggestions. The institutional level could take care of (meta)data quality control and invest in support for researchers (technical as well as in staff). Furthermore, open research data should be taken into account for evaluation, promotion and getting a doctoral degree. In case of funders, open access should be a requirement for funding and they could set up certain policy standards and mandates for researchers. The role of publishers has not been mentioned often, but publishers can help researchers to deal with research data, metadata and data storage.

4.9 Framework

The main focus of the framework needs to be, explicitly described by five experts: as open as possible, closed if needed. The **default should be openness** (mentioned as important in the second session by one expert), but sometimes it is not possible because of sensitive data (mentioned in section 4.2). Especially data from the social or medical sciences. One of the experts states that the open access to research data could be defined as ‘**intelligently open**’, implying that there are ways in

which data should be openly available and recognising that just making data available may not be enough (see also the aforementioned difficulty that data cannot be easily verified without context and provenance information). Another expert defines ‘intelligently open’ in the second session differently by referring to the entirety of the data available in a specific context and making an intelligent selection of what data are relevant and should be made accessible and what data may not even be worth storing for an extended period of time. There simply is too much data available and not everything can be stored “*ad perpetuum*”, maintained, or transferred to new information technology (IT) infrastructures whenever the previous one becomes outdated. According to Mr. Spichtinger (interview), the researchers know what kind of data are relevant for future research. However, another expert in the second session and Mr. Hof state that it is difficult to know what data are relevant in the future. Maybe data will be relevant for other researchers without knowing its usability in advance.

An important aspect of the framework, mentioned by one expert in the second session is that the framework must *not be too wide in scope*. Otherwise it will become either too general or too restrictive over too many domains.

4.9.1 Criteria

Criteria for open access to research data as guideline for a regulatory framework are:

- Data that is concerned to be **freely accessible** (following the line of the EC in its pilot in Horizon 2020 and the aforementioned point of view of what open access; see also section 4.2) is the:
 - *data underlying publications*, because this information content is already published to a large extent in the form of the publication;
 - *data, including associated metadata*, needed to validate the results presented in scientific publications;
 - *Other data, including associated metadata*, as specified and within the deadlines laid down in a mandatory ‘data management plan’; and
 - *Information about tools and instruments at the disposal of the beneficiaries* and necessary for *validating* the results (and – where possible – provide the tools and instruments themselves).
- The **approach of the European Commission in its open access data pilot in Horizon 2020 is appropriate**, according to the experts and Mr. Van Loenen (interview). All beneficiaries in actions that participate in the pilot must deposit all relevant data in a suitable repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and

disseminate it, free of charge for any user. The Horizon 2020 rules allow for exceptions in order to comply with confidentiality and security obligations or obligations to protect personal data. They allow exceptions where the provision of (open) access to some of the data concerned would jeopardize the objectives of the project. So, “the beneficiaries themselves determine to a large extent which data can be made accessible or not, but they need to explain why they include some data and not some other”;

- **Discipline/field dependent approach**, because of the aforementioned community which forms an environment with similar reasons and focus. Furthermore, “concerning the different disciplines, it is clear that research data in for example biomedical research has little in common with data collected in the context of research on medieval history. Whatever approach is taken by the different actors needs to take these disciplinary differences into account – there cannot be any “one size fits all” solution (also explicitly mentioned by Mr. Spichtinger, interview). Open access to research data should be discipline dependent, according to Mr. Hof (Interview). Criteria for that cannot be given by Mr. Hof, because it is difficult to oversee the relevance of data in the long term;
- **No strict formalizations**, a basic framework setting minimal standards is desired. Too strict formalization could hinder creative and practical implementation of the general ‘open movement’;
- **Incentives for sharing data and re-using data** (follows from the previous criterion, mentioned by two experts and Mr. Spichtinger, interview). Mandates are essential in this sense, because it ensures that everybody shares and consequently can benefit. Additionally, it will encourage researchers to invest in good data management. See for more information in section 4.10;
- **Taking into account the commercial interest of the business**. Otherwise businesses will not cooperate in (publicly funded) researchers because they will lose their competitive edge by giving access to their data (Van Loenen, interview). The government should listen to the business industry, according to Van Loenen (interview), “we have to entice business” to give access to their data (Van Loenen, interview). Cooperation between the government and the private sector is important for the well-being of the Dutch society;
This could break monopolies on open access to research data (see more information in section 4.8.1 and 5.1). A (governmental) framework may disturb competitive markets if it does not take into account the commercial interests of businesses (see also section 4.11);
- **Harmonisation of contradictions in open access definitions between funders and other stakeholders** are mandatory. Contradictions will result in a variety of approaches and interpretations, which create complexity for open access stakeholders. “Simplification efforts

must remain participant-focused, and should result in funding programmes that are easily understood and navigated, and cost-effectively administered by funders and participants and laid down in the guidelines for application and in the general agreements”;

- **No extra costs** involved for individuals who want to access the data. However, an expert in the second session questions who will pay for the data waiting to be used;
- Existence of a **data management plan (DMP)** for every project creating data (mentioned afore implicitly). Implementation of data management obligation are needed in the early stages of research. “This reduces the additional burden for researchers when meeting requirements for open access at the time of submitting a research proposal. A DMP may consist of an overview on the general types of data, data products and samples; a brief data description with respect to quantity and locations of data, a data analysis summary; radio buttons with respect to field work and use of existing data”. More in-depth information about a DMP will be given in section 4.9.2;
- **Specification of a licensing scheme.** An example for this is the Creative Commons Zero licensing (CC; mentioned by two experts). See also more information in the following section;
- Several practical technical (related) issues should be taken into account. First of all the storage and accessibility of data should be accessed through **trusted repositories**. Trusted repositories have a Data Seal of Approval and more information can be found in appendix F. Robust data infrastructures will optimise the way of dealing with data. Furthermore, the repositories needs to be **stable** and there should be agreement about the data models. Five-star open data is suggested by one of the experts and by Mr. Hof (interview): open licence, unique ID per object, non-proprietary formats, structured data and link(s) to other data), without using proprietary formats. In particular unique ID codes are suggested, also called **Persistent Identifiers (PI's)**. This is “an unique code that is associated with a digital object, whereby the object can always be found - also by changes in location and name” (Data Archiving and Networked Services, n.d.). An example of a PI is the **Digital Object Identifier (DOI)** and have been suggested by several experts and interviewees. A DOI “is an identifier (not a location) of an entity on digital networks” (DOI, 2013). It provides an identification system for exchange of managed information on digital networks.

Furthermore, data should be defined in **data types** and ‘categories of data’ could be made (for example what is meant with sensitive data and what does this constitute). A suggestion to ease implementation and communication of sensitive data, risk classes for privacy and commercial interest could be established. For example by using ranges from 0 (public domain) to III (very high risk). Overall, the data needs to be **found easily** and the location of the stored data should be found clearly by individuals inside and outside of

academia. Furthermore, the data should be **interoperable** (and machine-readable) through different computer systems and language(s) should be readable. Additionally, **standards** which have been set in a particular field of science should be **maintained**;

- **Decide if there is a necessity for establishing a period in which the data is available.** A period of ten years has been suggested by one expert. Examples from open access in astronomy, climate sciences and economics (where open data actually works in practice) suggest that such a limitation is counterproductive, according to one of the experts in the second session;
- **Environment of awareness, trust** (mentioned by five experts), **recognition and support.** Mandates and pressure will help to speed up procedures, but compliance will not follow if stakeholders are not convinced by the benefits of open access to research data. Trustworthy digital repositories are one of the aspects of this environment (see more information in appendix F and section 4.9.4). Therefore stakeholders have to work together towards a common goal by people who have interest in creating and maintaining a dataset (mentioned by two experts). Additionally, there needs to be a common ground in what open access can do and that the effort in making data available is worth the effort in terms of re-usability, actual use, institutional recognition and reciprocity (experts; Van Loenen, interview). In particular: “if a researcher spends a lot of effort making his or her data available and producing for example meaningful metadata and then nobody uses it, this will be not a good incentive to do it again”. Similarly, if an individual makes its data available and provide useful input to others’ work, but nobody else does so, this will not stimulate the open access process. Besides, cite-ability is essential, because trust, recognition and support depend on the way the researcher cites the datasets in publications, reports and on websites. Furthermore, cite-ability is essential for evaluating research data within the academic system (Spichtinger, interview; see more information on incentives for evaluating research data in section 4.10)

Cite-ability

The website of *DataCite* helps researchers to find, access and re-use data. It provide services which helps to find metadata, where DOIs can be tested and statistics of DOIs can be given. | www.datacite.org

- There should be **agreement on who is the owner of data and which responsibilities that ownership entails.** If possible concepts like ‘ownership’ should be defined. But as aforementioned, agreement on this topic does not yet exist. One of the experts defines it as an ambition. “Putting the cart before the horse, by starting to require a change to two

millennia legal tradition". The expert doubts whether this will succeed. See more information in the section before on how difficult it is to define ownership;

- "A general framework to decide **which values weigh heavier compared to other values** (for example: not always privacy as the most important deciding factor, but also take into account economic value, societal value, efficiency and effectively, practical concerns like technical (or even political) feasibility) while at the same time not being too strict to become too bureaucratic and impractical";
- **Researchers need to be trained properly** in data preservation and curation. In the second session one of the experts states that this is already the case at PhD level. Researchers do need to have time for research, according to one of the experts in the second session. In particular, researchers need to be able to obtain funding, stimulate innovation and do all sort of other things, this will be a lot together. Therefore, researchers need to be trained properly in data preservation and curation, because this saves time at the end of a research. Access can then be arranged easily;
- **Prolongation of mandatory clauses according to intellectual property (IP) dissemination and exploitation in EU-funding contracts.** A recommendation would be that future EU-funding contracts include mandatory clauses according to IP dissemination and exploitation". However, in order to have an effective policy, it is crucial to be highly specific with regard to a patenting-compatible schedule and/or the contribution percentage and company profiles in these mandatory clauses. In addition, according to one expert: "adopting open access calls for improvement and acceleration of the classical investment exploitation directive." Once the data are in the public domain, no royalties can be obtained for them nor can patents be obtained and therefore exclusivity of exploitation rights cannot be granted. Speeding up the process of intellectual property through an improved decision process can be beneficial to valorisation through classical investment exploitation. This process can be designed in such a way that it is not a significant hindrance to IP-access for industries. For example, the United States called for speeding up of the commercialisation process of the IP in the Open Data Policy, which in turn would be an improvement for the valorisation process in the universities and research institutes as such (mentioned by one expert; Project Open Data, n.d.); and
- **Sanctions** in case of non-compliance by researchers (mentioned by one expert). Another expert suggests that there should be a maximum deadline for making data accessible, like an embargo period.

One of the expert states in the second session that there is a necessity for standards and if it does not exist, new standards needs to be developed to ensure interoperability and sustainability.

Another expert in the second session suggests that the framework needs to be a robust ethical framework. This is used by researchers if necessary rather than developing new standards and therefore risk creating obstacles to future scientific developments. The expert states that an ethical framework is like a “community of humans” who value the data and want it to grow and prosper. Such a community can be set up around an existing journal and can be a virtual community, based on similar reasons and a way for sharing research data, to inform, check and inspire members or researchers in that community.

4.9.2 *Data management*

The data management plan (DMP) has been mentioned several times and more specification can be found in this section. Data management is essential in the process of open access to research data, according to the experts. A DMP indicates how research data can be administered, documented and shared (DANS, 2013b). It could entail an overview on the general types of data, data products and samples; a brief data description with respect to quantity and locations of data, a data analysis summary; radio buttons with respect to field work and use of existing data.

One of the criteria for a regulatory framework for open access to research data is a DMP. Therefore more specification is necessary of what is involved in order to meet appropriate criteria for open access to research data. One of the expert states in the second session that research data (raw) need to be defined. Does it refer to data management in general? Or does it refer to data management plans? If it refers to the first then funders of research and research societies need to pick that up. In the following textbox more information can be found about best practises of DMPs and insightful tips and reminders for establishing a DMP.

As mentioned before, as an aspect of the framework, there have to be sufficient incentives for researchers to prepare, store and publish their data. The data management plans should be implemented bottom up, according to one of the experts in the second session. “Learn from researchers what their needs are, because every discipline has its own standards, information processing and data collecting practices.”

Based on the knowledge of experts in the Delphi- analysis the following issues need (at least) to be taken into account for establishing a DMP:

- the period of how long the data should be kept;
- whether it is necessary to impose an embargo period, providing access only after the end of that embargo period (also mentioned by Mr. Hof, interview);

- how to ensure long-term storage and accessibility, including conversion to new formats if and when necessary, with a proper back-up system in place;
- privacy or security issues, and whether these could be approached by providing limited access (for example, in anonymised form, or only upon presentation of certain credentials);
- clarification of legal issues (ownership, including responsibilities and access rights for other partners involved in the same project, also beyond the duration of the project);
- authorship / who has contributed what (and should thus receive credit if the data are re-used / cited) – this information needs to be included in the metadata;
- create and sustain a community or ‘practice’ of open data (including incentives for sharing data);
- economic issues, by specifying efficiency and effectiveness; and
- where the data are stored (during and after research project). Data storage should be sustainable, which entails that it can be stored for a long period and maintained after publishing (Hof, interview).

Some aspects are not desired in the DMP, like categorization of data or access control since this would differ from discipline to discipline and may very well not be regulated at all (mentioned by one expert).

Mr. Grosfeld (interview) states that the DMP debate leads us away from the key issues in the open access to research data debate. First of all, clarity on definitions of open access, research data and raw data et cetera should be made. If these issues have been defined, there need to be agreement on who will finance the open access to research data and the maintenance of the data storage systems.

On the following page, best practices of DMP will be presented by giving brief information about organisations and its website.

Best practises

Several organisations made insightful overviews or list related to DMP and the implementation of it. or gave hints for researchers for establishing a DMP.

The Digital Curation Centre made a good overview of what needs to be considered for a DMP by enumerating questions and topics. Topics are the administration of data, data collection, metadata, ethics and legal compliance, and data storage, selection and sharing.

Digital Curation Centre | www.dcc.ac.uk

Publication | www.dcc.ac.uk/sites/default/files/documents/resource/DMP_Checklist_2013.pdf

The Dutch Data Archiving and Networked Services (DANS) organisation made a list of hints for researchers for establishing a DMP. It contains information about the preparation, implementation of a DMP, data archiving and drawing a DMP.

DANS | www.dans.knaw.nl/en

Publication | www.dans.knaw.nl/sites/default/files/file/Datamanagementplan%20NL%20def%20web.pdf

An example of excellent implementation of DMPs is the Max Planck Society, which focuses on basic research in the natural sciences, life sciences, social sciences, and the humanities. Another best practise is the Wellcome Trust organisation. This is a global charitable foundation dedicated in biomedical research and the medical humanities. Both organisations actively work with DMPs.

Max Planck Society | www.mpg.de/en

Wellcome Trust | www.wellcome.ac.uk

4.9.3 Data authenticity

A check of the data before publishing and the verification of data is another difficult issue (also mentioned by Mr. Van Loenen, interview). Ownership is again part of the debate of verifying data, because the owner of the data needs to decide who will check and verify the data before publishing. Because of the difficulty to define ownership, data verification also becomes difficult. Two experts state in the second session that the (main) data owner or rather the “registered subjects” need to check the data before publishing. Other experts also discussed the data verification and authentication issue by taking specific data cases.

In general, **data verification cannot be approached in the same way in each case**. Firstly, a quality check of the metadata, file format et cetera should be carried out by the repository, for example by archivists or depositor of the data (mentioned by four experts in the second session). This makes sure that the metadata contain all the relevant details, that the file is in a non-proprietary format, that any specific software that may be necessary to access and use the data is also included with the submission or is accessible elsewhere (with an indication where, and under what conditions). All this is of a technical nature rather than concerning the data itself.

Secondly, the intrinsic quality of the data is another issue. According to one expert in the second session the quality assurance should be done by peers, like it is the case with publications. Another expert is clarifying that the quality of the data itself must be ensured by the owner of the data (or the person that has been delegated as 'responsible' for the data. Quality of data in this context entails that the data have been gathered using a proper methodology, for example that it has not been falsified and that it is relevant.

Thirdly, in the case of data underlying a publication, these data should be made available to the peer reviewers of the publication, who may then be able to spot any 'issues' of the data as well (for example, if it is apparent that the data has been manipulated; mentioned by two experts in the second session). However, this may burden the review system even more. It is a serious issue, which should be taken into account, according to one expert in the second session.

4.9.4 *Data storage*

Data needs to be preserved in sustainable, certified data repositories, according to the experts and Hof (interview). The information and data should be stored in such a way that it is easily **discoverable** and **accessible** and where this is ensured or can reasonably be expected for a long time. This could be institutional repositories if (and only if) these are interoperable with larger networks and linked in such a way that the content can easily be found via this larger network (mentioned by one expert in the second session). Researchers do not know the possibilities of storing data, this should be more known (Hof, interview). Another issue in data storage is that there is no (or not enough) budget for data storage. Often data should be transformed in other formats in order to store it online (Hof, interview; Kolman, interview). Frequently, there is no money for transforming the data because it has not been calculated in the project budget. Besides money, data storage can often not be done at a regular server. Additionally, data should still be accessible within ten or fifty years (Hof, interview; Kolman, interview; Van Loenen, interview) and an organisation should be responsible for that (Hof, interview).

Metadata, as mentioned before, is very important for data storage. Publishers can help with the infrastructure of data storage because they have experience with data storage (Kolman,

interview). Linking datasets and data with articles is well-known. Mr. Kolman, working for the publisher Elsevier, describes that they as publisher can help researchers to deal with research data and how to describe data (metadata). Furthermore, they could help in digitalising data and informing about data storage and discoverability. In particular, data are often not recorded properly and data are often only stored at a private computer, therefore researchers need to have help in recording data and data storage, according to Mr. Kolman (interview). Additionally, research data are often not digital (Van Loenen, interview). Elsevier can enrich raw data and transfer data (Kolman, interview). Good ontologies (in short: data structure that indicate types and interrelationships of data) are necessary (Kolman, interview). People should not underestimate how complex the data storage is, because aspects as well-defining data, proper metadata, discoverability and long-term storage are difficult to establish, according to Kolman (interview). Furthermore, new soft- and hardware complicates data storage. Publishers could play a role in data storage and maintenance, Elsevier is willing to help with that, according to Mr. Kolman (interview).

Research data can be stored on “either a national level or international level depending on several things”, according to one of the experts in the second session. Even governmental registries would be an option, whenever it is possible and feasible.

According to one expert in the third session, in many cases it will make more sense to deposit the data directly in subject-specific repositories that are maintained in a professional manner, with the necessary expertise to deal with for example data formats that are specific to the subject area in question. Again, these repositories should be interoperable with relevant larger networks and ensure discoverability via those other networks.

For very large data coming out of big projects, the best solution may be to set up a repository just for those data only. This will be especially relevant for multimillion Euro projects involving a large number of partners and running for a long time. But here again, discoverability must be absolutely ensured, for example by making sure that the repository is listed in relevant directories, according to one of the experts in the second session.

At the end of this section a textbox of mentioned websites of organisations can be found. Repositories as Dryad can be appropriate for (small) data underlying publications which allow easy linking to the publication itself. This would also allow the allocation of an individual persistent identifier (DOI or similar) to the data, thus allowing individual citation of the data. Another experts states that *individual datasets* from small-scale experiments that have led to a publication, and if this dataset will be necessary to verify the results, then it is most appropriate point that this dataset can be linked to the publication. If the research is *local* it is most likely to use local (national) services. In this case an *institutional repository* that provides persistent identifiers and that provides reasonable

assurance of sustainability would be an acceptable solution. In this case the aspect of discoverability is not so important as it is rather unlikely that somebody who is not familiar with the publication will be looking for the data from the given set of experiments.

Depositing the data independently to a repository and linking it to the publication would therefore be preferable to simply providing the data to the publisher as additional files. Provided that the repository is trustworthy and has a proper back-up system in place (an example is CLOCKSS), this would also secure the accessibility of the data independently of any publisher access policies. Furthermore, storage during research should be done locally, by example SharePoint of Microsoft or the Dataverse Network (according to one of the experts).

Long-term storage of data should be taken care of by a network of 'Trusted Digital Repositories' which aim is "to provide reliable, long-term access to managed digital resources to its designated community, now and in the future" (DANS, 2013a; Research Libraries Group, 2002, p.5). In addition, if there will be talked about *large scale observations* or *experiments* that produce huge amounts of data that will offer opportunities for analysis for many other researchers, then discoverability is of major importance. In this case *national* or even *international* repositories that cater to the specific scientific community concerned (and are interlinked with other similar repositories) will be the more suitable ones as they will ensure much more visibility are discoverability to the data.

There are examples of national institutional repositories, like DANS-KNAW in the Netherlands, which offer practical solutions for (specific) data preservation and storage (mentioned by three experts in the second session). Furthermore, research data from projects in Horizon 2020 could be a stored differently. Ideally this would be stored in a transnational repository, according to one of the experts. More and more private actors seem to deliver services of hosting data and public organisations also make use of it. For example the aforementioned tool of Microsoft.

One of the experts in the second session states that the specific 'location' or manner of storage is not a fundamental concern, the central idea is to be **pragmatic**.

In the third session experts commented on criteria to choose the right and appropriate data storage. A general principle for the choice of an appropriate repository should always be that it **should fulfil its purpose**. Furthermore, the choice will be dependent on the availability of storage that will preserve data sustainable, make them easy discoverable, cost-saving and accessible (according to two experts). The related "technology is a means to achieve an end, it should not be leading". Essentially, discipline specific repositories need to be set up, managed by stakeholders who are knowledgeable about the subject, both to ensure continued relevance and quality of added contents (for example the holder/processor; mentioned by three experts). Only the communities who value

the data (other than as a pay check or because of the open access as a principle) will be able to offer a sustainable community in the long run. However, there is a difference between an open access repository and a digital, deep archive. The first is easier to have in each institution. It is relatively cheap and institutions can showcase their research. The latter is far more challenging at institutional level.

Additionally, there are several certification systems that cover the most important criteria for good repositories. The Data Seal of Approval (DSA) provides assurance that at least minimum standards are fulfilled by the repository, as explained in the DSA Guidelines in appendix F. There are other more far-reaching certification mechanisms as NEN or ISO norms which are managed by experts, such as DIN 31644 or ISO 16363 certification (DIN, 2014; ISO, n.d.; mentioned by two experts). All repositories should strive to fulfil as many of these criteria as possible. Institutions and funders should encourage the use of such repositories and provide incentives for doing so (and perhaps in the long run – but not immediately – also make it mandatory).

Examples of data storage

If institutes have enough of the same type of data they can do it themselves. For example, Naturalis in the Netherlands gives access to the database of botanical specimens in their collection. The European Bioinformatics Institute (EBI) holds genome and other data that is provided by (and can be searched by) researchers all over the world. The connected European research facility ELIXIR.ERIC is going to do the same for other molecular information, but in a distributed manner.

Naturalis		vstbol.leidenuniv.nl
European Bioinformatics Institute		www.ebi.ac.uk
ELIXIR.ERIC		www.elixir-europe.org

Other examples of repositories are:

(Small) data underlying publications		Dryad		www.datadryad.org
Trustworthy and proper back-up		CLOCKSS		www.clockss.org
Storage during a research (local)		SharePoint		office.microsoft.com/en-gb/sharepoint/
Storage during a research		Dataverse Network		www.thedata.org

4.9.5 Finance and maintenance

Obviously, data storage needs **financial support** and should be **maintained** (suggested by experts and Hof, interview). This difficulty has been mentioned by two experts in the second session by questioning how more specialised or disciplinary repositories may kept sustainable. Since in most

cases national repositories are financed by research money (or government money). Publishers have their own repositories and for the commercial publishers new business cases could evolve, but could also affect the openness of data immediately. This limitation of open access to research data has been recognized by two experts in the second session. Additionally, one expert states that in the long run it is probably not possible to have a certified data repository, a persistent one, in every institution. Therefore, this expert states that European member states could play a role in this process.

Another expert in the second session states that the financing of a database should be the responsibility of the public organisation or research institute concerned with gathering the data in the first place. Nevertheless, co-financing may take place to make these systems accessible to a more general public.

According to another expert, funding of a data storage system depends on the kind of repository or database. If it is an institutional repository it should be financed in the same way that any other infrastructure at the institution is financed (mentioned by one expert), through allocation of indirect costs under research grants, from core funding, or similar. If it is a repository that has been created specifically for a project, then it should be financed from the project funds while the project is running, and then from other institutional funds. In both aforementioned cases the repository or database should be maintained by specialised staff at the institution (data scientists) who do this either as part of their normal job or who have been hired specifically for the purpose on project funding. In any case, it should *not* be the researchers themselves who are burdened with the maintenance of the system.

If it is an external database or repository, then different business models are possible. It could be funded by fees that are payable when a dataset is submitted (this is the case for Dryad), similar to the Gold Open Access fees for publications. Or the repository could be sponsored by third parties. However, according to one expert in the second session here it would be preferable if the sponsor were a not-for-profit entity rather than a commercial entity that may at some point try to make profit from the database by restricting access to the data that has been deposited unless an access fee is paid (thus defying open access).

Maintenance would, in the long term, have to be taken care of mostly by the community of (re)users, who share this responsibility with the data gathering organisation (mentioned by one expert in the second session). Furthermore, DANS (2013a) made a document where the responsibilities of various stakeholders in the open access process have been defined:

The basic technical infrastructure is provided by data centres, an area where parties like SURFsara and Target have a coordinating role on a national or regional level. The back-office functions are carried out by organisations with a national role to play in the field of long-term

accessibility of data in trusted digital repositories, such as DANS and 3TU.Datacentrum in the Netherlands (red.), collaborating in Research Data Netherlands. Together they have expertise on data from the humanities, sciences and social studies. The front offices are located at universities (libraries, local data centres), research/knowledge institutes, colleges of applied science, national and international research infrastructures (ESFRI/National Roadmap); for some features they can also be found with the funders (Dutch organisation as NWO, ZonMw and Ministries). All these organisations have in common that they are primarily responsible for the quality assurance of the data produced and processed by them or for them (DANS, 2013a, p. 4).

4.9.6 Licensing scheme

As mentioned before, more clarity about a licensing scheme is necessary. In general, the licensing scheme should be as open as possible according to five experts in the second session. Low threshold scheme would be most appropriate. Four experts explicitly suggested a license and all those four experts and Mr. Spichtinger (interview) suggested a Creative Commons license, specifically CC0 (no rights reserved), or CC-BY 4.0. The latter license allows to share and to adapt the material for any purpose, even commercially. The DANS institute uses a data license agreement between the repository/database and the person entitled to the dataset (DANS, 2013). This can be a person or an organization. Such an agreement could be a suggestion for other stakeholders in the open research data debate.

In the following textbox more information can be found about data licensing.

Data licensing

A licensing scheme is necessary and the Creative Commons (CC) license has been mentioned often in this study as a low threshold scheme. In the text below two CC-licenses have been described.

CC0 license | No rights reserved

“CC0 enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law. In contrast to CC’s licenses that allow copyright holders to choose from a range of permissions while retaining their copyright, CC0 empowers yet another choice altogether – the choice to opt out of copyright and database protection, and the exclusive rights automatically granted to creators – the ‘no rights reserved’ alternative to our licenses.”

CC-BY 4.0 license

CC-BY 4.0 enables sharing by copying and redistributing the material in any medium or format. Furthermore it enables adaptation by remixing, transforming, and building upon the material for any purpose, even commercially. “The licensor cannot revoke these freedoms as long as you follow the license terms.”

Licensing agreement

Data Archiving and Networked Services (**DANS**) uses a data licensing agreement (in Dutch):

www.dans.knaw.nl/content/dans-licentieovereenkomst-gedeponeerde-data

www.dans.knaw.nl/sites/default/files/file/EASY/Licentieovereenkomst%20DANS_NL_4_2.pdf

4.9.7 Protecting data

The method to embed the security of data is a difficult issue. Out of six experts in the second session who responded to this question, only one expert gave an opinion on the security of data. According to that expert the security of data concerns two issues:

1. **Access control**; making sure that sensitive data are only accessible to those that are entitled to it; and
2. **Keeping the data safe** in the sense of keeping it accessible and usable; This entails the upgrade to new data formats and updating of software that may be necessary to access or use the data, setting up reliable back-up system, et cetera.

“Access control is not a trivial issue, especially since the security measures have to be constantly upgraded in order to respond to technological advancements opening up new possibilities to hackers. But also the second aspect, making sure that the data remains accessible and usable, can be a challenging task. Small institutions that may not have the necessary know-how should therefore consider very carefully whether they really want to run their own repository or whether it would not make more sense to team-up with specialised data centres that would host their data in a secure way”, according to one expert in the second session.

4.9.8 *Framework feasibility*

The framework could be implemented according to two experts, but it would require **willingness** and real **commitment**, in particular a monetary commitment. As mentioned before, “data storage and management and everything that is related to it (starting from the training of young researchers to become more data aware, to the support for researchers in preparing their data for deposit, such as preparation of metadata, the hardware necessary and the technical support and maintenance) costs money. Given the current economic climate (especially in some European countries) it will require a strong political commitment to support the allocation of funds to these areas, while funding is cut in other areas that may seem more relevant and of more immediate benefit to some stakeholder groups.”, according to one expert in the third session.

In addition, it will also be crucial to have a mechanism for continuous **open dialogue** between the government and the different stakeholders concerned (mentioned explicitly by one expert). In this way there the buy-in and continuous feedback from those that are directly affected by the measures proposed and implemented can be ensured. Stakeholder include (at least) researchers, research institutions, research funders, professional societies and publishers. The expert gave an example of the UK Open Research Data Forum, which is an existing initiative for an open dialogue (UK Open Research, 2014). Working together and having respect for each other’s role and possibility is essential. Furthermore, forcing people to do things will not help much. Therefore incentives like giving credits for preparing data and making it available and training will be much more effective. People need to see and recognise the benefits themselves. So it is a delicate framework of mandatories and soft laws.

Another critical issue in open data according to one of the experts in the third session, is to start small, in one domain and then expand. Introducing both a broad range of responsibilities and roles (as here) over many domains at once, it will most likely fail. In addition, the **step by step** approach will allow more flexibility, which obviates the need to be complete upfront, a practical impossibility for the future anyway. Furthermore, do not collect everything, a system should be set

up to make decisions on what to keep and what to delete. In the end it is not an easy process and the proposed framework will be feasible, but it will take time.

Summary | All experts of the Delphi-analysis do agree on the necessity of a framework for open data, as well as several interviewees. The focus should be: ‘as open as possible, closed if needed’. The default should be openness, but sometimes it is not possible because the data are sensitive. The term ‘intelligently open’ could therefore be appropriate in this context. According to experts, data that should be freely accessible are: data underlying publications, data needed for validating results, metadata, data which have been specified in an agreement and information about tools at the disposal of the beneficiaries. The approach of the European Commission in its open access data pilot in Horizon 2020 is appropriate and a final framework should be discipline dependent. A framework should harmonise contradictions in open access between funders and the definitions in use and there should be no extra costs and strict formalizations. Commercial interest of businesses should be taken into account. Furthermore, a decision should be made on the necessity of an embargo period (mentioned by three experts and Mr. Hof, interview). For researchers the framework should create an environment of awareness, trust, recognition and support. Ownership should be defined and the researchers should be trained properly. Additionally, mandatory clauses according to intellectual property dissemination and exploitation in EU-funding contracts should be maintained. Possibly, sanctions can be implemented if there is no open access to research data.

An essential aspect is a Data Management Plan (DMP), additional with specification of a licensing scheme (as Creative Commons) and recommendations for trusted repositories. More specification of what is involved is necessary in order to meet appropriate open access to research data. Clarity on data authenticity (control) and data storage (national, institutional or discipline dependent) should also be part of the framework. Overall, data management should have finance and maintenance (which depends on the kind of repository or data storage). Protecting data should also be one of the aims in data maintenance. The feasibility of this framework depends on willingness and commitment of all stakeholders and it will be crucial to have a mechanism for continuous open dialogue between the government and the different stakeholders concerned. A step by step approach, probably even starting with one domain, is suggested for this complex process and it will take time to implement open access to research data in general.

4.10 *Incentives for researchers*

Incentives are also important to convince researchers to give open access to their research data. The framework described above has mentioned the demand for incentives and several interviewees also mentioned it (Mr. van den Biesen; Mr. Kolman; Mr. Hof; Mr. Grosfeld, interview). Academic careers

should be helped and not hindered, according to one of the experts in the second session. **Therefore one of the key requirements is to recognise data sharing as publications** (mentioned by two experts and Mr. Hof, interview). Three experts are also of the opinion that the efforts researchers put into the creation, management and publication of data part of the system of academic credits. This accreditation of their effort will work out quite well. Additionally, “large chunks of data could still be scooped up by a competitor leaving the smaller less well equipped originator out in the cold, this is foreseeable with weaker countries in the EU as well”. Currently, ‘**data articles**’ can be published, which is more an article of metadata (Hof, interview). This is an article which describes what data have been used and where it have been stored. Furthermore, background information about the owner and methods will be described in a data article. Several journals will accept these kind of articles, in particular peer review journals. By doing this, researchers can receive credits and they could add it on their publications list (Hof, interview).

One of the experts suggest that either the culture needs to change (which is not an easy task), or there has to be some form of protection for the originator to enable him/her to use the data also to further his or her career (for example by determining who can get access). Additionally, **re-use of data needs to be encouraged for example by funders and within the academic system**. “Support needs to be provided by universities and research institutions, including direct support to researcher in the curation of their data, but also by including relevant skills training in the training programme of young researchers (possibly with funding from national funding bodies)”. In addition, researchers will give access to their data if it is an advantage for them as researcher or their career (Spichtinger, interview). At this moment, this is not the case because researchers are evaluated on how many publications you have in a high evaluated journal. This is probably an old-fashioned approach (Spichtinger, interview). The University of Liège is evaluating researchers on all the articles they put in their repository (Spichtinger, interview). In this way researchers are encouraged to upload their research data and to give access to data. This could be an example of implementing an incentive for researchers within the academic system.

Other (related) incentives for researchers to give open access to their data are:

- An emphasis **on improvement of efficiency and effectiveness**, which heightens research integrity, trust and legitimacy. Furthermore, emphasising the benefits that come with forming a vibrant research community;
- Activities related to data management and data sharing must be **acknowledged within the career system**, for example, for promotions and other forms of career advancement. This will require a fundamental reform of the whole academic system. They should also be recognized by funders when assessing the scientific merit of individual applicants for research funds;

- **Funding** schemes should include research data management and sharing and the associated support in an integrated way – proposals that present a well thought-through data management strategy and provisions for proper data storage beyond the duration of the project should receive credit for this in the evaluation, and costs associated to these activities should be eligible, also for a reasonable period after the lifetime of the project if this is justified;
- **Common standards for citation** of data must be established, together with a proper (internationally accepted) mechanism for indicating what the contribution of the individual is to the data concerned. Suggestions for a common standard for citation of data has been established by Force11 (FORCE11, 2014);
- Include the **requirement for open access to research data in evaluation criteria** for tenure, projects, PhD admission, for example; and
- **Invest in altmetrics.** Altmetrics is “the creation and study of new metrics based on the Social Web for analyzing and informing scholarship.” (Altmetrics, n.d.; Priem, Taraborelli, Groth & Neylon, 2010). It is about creating and establishing metrics for analysing the reach and impact of published research (PLOS, n.d.2). When a reliable base exists, these metrics can be used for evaluation. The altmetrics movement describes: “we rely on filters to make sense of the scholarly literature, but the narrow, traditional filters are being swamped. However, the growth of new, online scholarly tools allows us to make new filters; these altmetrics reflect the broad, rapid impact of scholarship in this burgeoning ecosystem. We call for more tools and research based on altmetrics” (Altmetrics, n.d.). Because of the demand for new metrics one experts suggests more investment in altmetrics.

Summary | Incentives are important to convince researchers to give open access to their research data. The foremost argument is that data sharing should be recognised as a publication, so that researchers can receive credits for their effort and it will get acknowledged within the career system. Re-using data should be encouraged by funders and within the academic system. Furthermore, funding (for sharing and altmetrics) and common standards for citation are necessary. External requirements for open access to research data should be mandated via evaluation criteria and could also stimulate open data.

4.11 *Business and industry*

The industry has been mentioned before several times, because the experts state that open access creates opportunities for business by giving access to support innovation. Nevertheless, the DMP needs to take into account the commercial interest of business. Therefore the experts in the second

session discussed the issue how to involve the private sector and industry in order to convince them to give open access to their research data and to share their data. Especially publicly funded research could limit the participation of business in projects of open access to research data. As mentioned before, businesses and enterprises are often involved in research what is partly governmental funded. If those companies need to provide their research data to public, then this could harm their competitive edge. Consequently, companies would probably be less interested to cooperate in publicly funded research. This could damage the research industry and therefore the interests of businesses and industry should be taken into account. Therefore, as mentioned before, the opt-out possibility in the EC pilot on open access to research data is important for businesses and industry to decide on a case-by-case basis whether research data will be accessible (Spichtinger, interview).

One of the experts states that many private companies do give open access to research data after a patent has been granted, but no examples are given in this sense. If companies cannot apply for a patent, then it will be difficult for private parties to participate in (publicly funded) research. These companies cannot make profit out of it, neither protect their activities and products. Another expert adds that it almost impossible for private parties to publicise their data without harming their commercial interest. In some domains data sharing would not harm the interests, or data sharing would lead to an increase of the benefits for most stakeholders. Only one or two stakeholders would probably not benefit from sharing data. Therefore the government can subsidise these stakeholders via a public-private partnership construction.

A way to convince the private sector is to point out the commercialisation aspects of open data. Data, which is licensed under a CC-BY license can easily be re-used. For instance app-developers can make use of research data like the human genome database, or cultural heritage, which is being digitized on a fast pace. However, big companies are not so easy to convince to make their data publicly available. Trade secrets or business models are at stake. An embargo period for opening up the research data and giving the companies enough time to earn their investments back would be a suggestion to convince private parties to open up their research data, according to one of the experts and Mr. Spichtinger (interview). Convincing the private sector and industry has been described by another expert by suggesting a mix between carrot and stick (as mentioned in chapter 2). “The stick would be that they will not be able to benefit from public funding or subsidies if they do not comply with basic principles of openness”. However, enough flexibility must be provided to take into account the specific interests of private entities. “The carrot could be additional (financial and other kind of) support for activities related to data curation and storage, including security measures. This could consist of free training sessions for the up-skilling of personnel, organisation of 'mutual learning seminars' and workshops for the exchange of best practice and to discuss common challenges and

solutions, (co-) funding for the necessary IT infrastructure, prizes for particularly open approaches et cetera. According to the expert, the concrete type of measure will depend very much on the type of business and the sector, but also on the national legal and political framework and 'business culture'. For example, an approach that works with high-potential SMEs in the pharmacy business will not necessarily work with a multinational insurance company.

4.11.1 *Philips perspective*

This business perspective in this section has been developed by Philips' Research department through an interview with the Vice President Public Research and Development Programs. The Philips position can be found in appendix E. A communication from BusinessEurope has also been used which describes the position of industry on open access to research data, see appendix D. These positions are combined in order to formulate a general position of the business industry on open access to research data. There is no input from the Delphi-analysis or other interviewees used in this chapter. However, in some cases the argument will be supported by mentioning Mr. Grosfeld from VNO-NCW – an employers' organisation in the Netherlands – as reference. The position on open access to research data of Philips and VNO-NCW overlap each other and certain statements can therefore be supported by mentioning Mr. Grosfeld as reference.

The industry sees open access to research data on the one hand as a threat, but on the other hand also as an opportunity, because it may give access to (new) information in the context of 'big data' and its applications. Data can be for example useful by extracting data to other disciplines or by combining information out of data sets. In spite of the threats described below, open access to research data could be very useful to companies. Sharing data could initiate new collaborations between companies.

According to Mr. Van den Biesen (interview), open access to research data should not entail that all data from publicly funded projects need to be publicly accessible. He states that it is unclear at the moment what is meant by publicly funded research (also mentioned by Mr. Grosfeld, interview); is it only when it is fully funded by the government or also when research is partly funded by the government and partly by business? "If all research data which is fully funded by the government, then the business industry does could not have any objection against it", according to Mr. Van den Biesen (interview). He adds that politicians are too naive by saying that public money for research comes and that research data need to be publicly accessible. This structure would limit the possibility to apply for a patent, according to Mr. Van den Biesen (interview). In particular, the protection of innovative products will be limited. These data could not be protected, because it is open to everyone and can therefore not lead to profit for companies. Summarising, industry will be

reluctant in investing in commercializing the results from research which is fully funded by the government, according to Mr. Van den Biesen (interview).

Consequently, research data in the aforementioned situation will not be used and will be useless in this context, according to Mr. Van den Biesen (interview). This is in contrast with the vision and aim of open access that all data which is accessible will be used. It rather will not be used because it is usable to everyone. So by widely spreading the research data it could destroy value rather than creating value. Consequences for companies will arise, but also for universities because of its valorisation function. The main motivation for companies to invest in research and developing a product is the possibility of protecting the results so as to ensure a competitive edge. Intellectual property is essential in participating in research collaborations.

The industry wants to decide for each project on a case-by-case basis whether or not the consortium can give open access to its research data (Van den Biesen, interview; Grosfeld, interview). Additionally, more clarity of *which* data needs to be made accessible is also a prerequisite of the industry. Data underlying publications would in many cases be no problem. However data from lab journals of researchers would be a “no go” for Philips and probably for the entire industry (Van den Biesen, interview). The consequence would be that (internal) ideas and notes must be accessible to the public, which hinders the application for patents by companies.

Additionally, industry wants to have the option for an opt-out from any obligation of open access to research data, according to Mr. Van den Biesen (interview). This possibility to decide in each case if the data will get openly accessible will give industry more leeway to carry out its activities. Giving always open access to data in certain situations could harm their position too.

Moreover, if open access is about raw data, then it is not clear what metadata entails (Van den Biesen, interview). More clarity is also necessary about this topic. Owning data and its analysis is a business model, so they may not give open access to their data.

In the contemporary European research programme Horizon 2020 the ‘rules for participation’ (specified in Article 43.2) take into account the interest of industry in the debate of open access to research data (European Commission, 2013e).

“Subject to any restrictions due to the protection of intellectual property, security rules or legitimate interests, each participant shall through appropriate means disseminate the results it owns as soon as possible”.

Consequently, companies or research organisations can opt-out in certain cases and do not need to give open access to their data, mentioned by Mr. Van den Biesen (interview). In the ‘model grant agreement’, the possibility of an opt-out has also been described, but not very clearly (Van den Biesen, interview). In the guidelines of the pilot for open access to research data the opt-out has

been mentioned obviously, but this document does have less legal force. The industry in particular wants to benefit from the data from the research that it has invested in. However, the Horizon 2020 pilot is very welcomed, but many aspects are still unknown, according to Mr. Van den Biesen (interview). Specifically, who is going to maintain a data system; who is taking care if the data is really authentic (without alterations before publishing) and who is going to pay the data (maintenance) system? Additional questions from the industry are for example questions as how will liabilities and responsibilities be framed and implemented in the open access to research data process? Which disclaimers will be needed? As a researcher, I suggested that the industry could (partly) pay for the establishment and maintenance of the data system (this was one of the suggestions in the first session of the Delphi-analysis). Mr. Van den Biesen states that it will be difficult to quantify the benefit out of the data. So, another question arises: who would pay for what and to what extent? Therefore the pilot is a good inventory for further (final) implementation of the open access to research data, according to Mr. Van den Biesen (interview).

To come back to the issue of publicly funded research, the definition of it is unclear (Van den Biesen, interview; Grosfeld, interview). Because where is the debate about? Is the open access to research data only for fully publicly funded research or also for research which has been publicly funded for sixty or even for twenty percent? Using a certain percentage for drawing a line in which publicly funded research projects needs to give access to their data. For example, the guideline could be that each research project which is publicly funded for forty or more percent of the full cost needs to give open access to their data. Research projects which are then publicly funded less than forty percent do not have the obligation to give access to their data. Philips states that it will be difficult to draw a line in the percentage of publicly funded data which needs to be open mandatory, and in any case fourthly percent of the full cost would be far too low. Philips doubts whether this approach would work (Mr. Grosfeld mentioned the same during the interview).

A remark on the aforementioned statement is that Philips does not have any experience yet with the open access to research data pilot, because of the short time that Horizon 2020 is in effect. Philips therefore opted a restrained approach towards the pilot of open access to research data and their research projects. However, encouragement could help the business to participate, according to Mr. Van den Biesen (interview).

In order to stimulate participation of the industry in open access to research data it is essential that there is an added value for sharing data in external consortia (Van den Biesen, interview). Reciprocity is also necessary for having a balanced partnership, according to Mr. Van den Biesen (interview). However, reciprocity assumes two parties instead of open access which means open to the public and not a specific stakeholder. One remark is important here, because enforcing reciprocity is also

highly difficult which complicates the open access process even more. Another way of persuading industry to participate in open data is using specific terms after which the data needs to be available. This would give room to apply for patents and further development. For businesses it should not be made mandatory. Businesses want to have the possibility to opt-out, if that is not possible this would have negative consequences, according to Mr. Van den Biesen (interview). This would have consequences for the whole society, because companies will probably participate less in public research. This affects the economy and consequently the knowledge and research institutions itself (for example universities). Moreover, valorisation and start-ups of new companies could be less frequently set up. So, the question arises: “do we really want to give access to our data?”

All in all, industry sees open access to research data as a threat and as an opportunity, but would want to be able to decide on a case-by-case basis. The more open access is made mandatory, the less attractive it will be for the business industry. **If all data always need to be open to everyone, business industry will face severe problems.**

Summary | The business and industry sees open access to research data on the one hand as a threat, but on the other hand also as an opportunity. They need to be convinced and unclear aspects need to be removed by for example answering questions as: what is research data and which data should be open? Especially, what is publicly funded research? The industry wants to decide for each project on a case-by-case basis whether or not the research consortium can give open access to its research data, an opting-out possibility is desirable. Opening all data to everyone will be problematic.

4.12 *International perspective*

At the European level, the European Commission (EC) established a pilot within Horizon 2020 which aims to improve the accessibility to research data. The EC gave additional information about the guidelines on open access (European Commission, 2013b) and guidelines on data management (European Commission, 2013d). In section 2.3 information about the European Union has been presented, this section gives additional in-depth information.

One of the experts gave a summary of the pilot of Horizon 2020. “The current wording of the relevant article in the Model Grant Agreement provides a lot of freedom for the beneficiary to decide about the data to be included, recognizing that it is the researchers themselves who ultimately are the best placed to decide what data can be shared and what not. At the same time, the requirement of a data management plan means that beneficiaries (and to a lesser extent already applicants) have to think about the topic at an early stage, and consider data sharing aspects in the planning of their research. So this is a kind of soft pressure forcing beneficiaries to engage with the topic, while not imposing any very strict rules on them”.

“It remains to be seen whether the current set-up will be successful. The pilot currently covers only a small part Horizon 2020, with crucial parts (such as the Marie Skłodowska-Curie Actions, the ERC programme and the programme on Health and Demographic change) not taking part in any structured way. Participation in the pilot on an individual basis is possible for any project under Horizon 2020 except for the ERC, but it is questionable whether many projects will sign up to the pilot without any clear incentives to do so. ERC projects from the 2014 Work Programme did not even have the possibility join the pilot on an individual basis; this will change with the Work Programme 2015. Participation by the Marie Skłodowska-Curie Actions in the pilot would have been particularly desirable as the programme trains many young researchers at PhD and postdoc level, and could have had a very useful structural impact on the training of young researchers on data related issues much more generally through important spill-over effects”. This has been considered as a lost chance by one of the experts, suggesting that it was “to a large extent also due to the fact that the relevant Unit in the Commission was not actively involved in the discussions.” In addition, the evaluation of the pilot will probably follow in a couple of years, which takes a lot of time (Van Loenen, interview). The first results of the pilot can be expected at the end of 2014, at least for some disciplines.

“Another crucial point to mention is the fact that the Model Grant Agreement foresees sanctions in case of failure to comply with any of its provisions (including the article related to the open research data pilot) and this in a much more visible way than was the case in FP7. However, it is highly questionable whether the Commission will have the capacity (in terms of human resources) or the political will to enforce any of those sanctions in the context of the open research data pilot (or in the context of open access more general, for that matter)”.

The awareness differs between the European member states considerably. The EC and some countries like the United Kingdom (mentioned by six experts) and the Netherlands are putting better data management on top of their priority list and increase awareness of the benefits of open research data and to provide support. This is to be achieved through training initiatives or the running of specialised data centres. The Netherlands wants to be a frontrunner in open access to publications and research data and places this issue high on the agenda of its presidency of the Council of the European Union. The European context is highly important for further steps in the process of open access, according to Van Loenen (interview). Other European member states do not yet see an immediate solution or do not exactly know yet where the debate is about (mentioned in the first session and modified in the second session). Also the attitude with respect to open access is very different between countries. Furthermore, many differences exist between repositories,

institutions and disciplines. Some are very much advanced with defined rules and agreements, others are in a starting phase.

“However increasingly there is a push for research data to be shared more widely and throughout the research process. But it is recognised that this is harder to mandate and also there are cultural barriers to this which are even harder to overcome than that of sharing data ‘behind the graph’ – behind the research results in a paper.”

The United Kingdom (UK) has set the standard in recent years. For example the UK research excellence framework (REF) has not yet sought to treat data the same as papers, but it is increasingly seen as important to make data accessible (mentioned by one expert in the second session; REF, n.d.). The policy for the next REF encourages data to be accessible by stating that credits will be given where it is in the research environment aspect of its evaluation of research. However, the open access to research data efforts seems to be mostly driven (and lived) by a small group. According to one of the experts, “it is interesting to review how Open Source communities in general have spread over the last (two) decades. Those communities share many similarities, including the hotly debated aspect of principles versus practice. Some of those communities have indeed become (and remained) thriving, others either did not or are no longer successful. Especially the latter category is interesting as they may give an insight in the type of communities which do succeed and the key ingredients of success.” There is also cooperation on open access to research data between business via Business Europe (Grosfeld, interview; Van den Biesen, interview). Commissions discuss the perspectives of businesses and together they work as a lobby organisation on open data.

More information about open access to research data in the United Kingdom and the Netherlands can be found in the following textboxes.

The Netherlands

In the Netherlands several institutions have been established that work on open access to research data. Several institutions and initiatives have been discussed during this study.

Examples of these institutions are:

Data Archiving Networking Services (DANS)	www.dans.knaw.nl/en
3TU.Datacentrum data storage and knowledge	http://datacentrum.3tu.nl/en/home/
The Language Archive data storage	https://tla.mpi.nl/

United Kingdom

The government of the United Kingdom made a approach to access and publish research findings. It contains a general approach of science which is “a spirit of free - and occasionally eccentric - intellectual enquiry”. The government recognises the value which publishers add. “Peer review is a crucial part of the research process. It takes various forms in which academics generously give their time to scrutinise draft articles. But value can also be added by identifying the academics to conduct peer review, through the editorial function of signalling which research is of the highest worth, and by helping others to find it.”

UK government | www.gov.uk/government/speeches/public-access-to-publicly-funded-research

TIP! | **Digital Curation Centre (DCC)** | DMP management tool by helping to write a DMP | see also textbox in section 4.9.1 for more information about DCC | <https://dmponline.dcc.ac.uk/>

Several institutions made policies and/or principles to establish open access to research data:

Biotechnology and Biological Sciences Research Council (BBSRC) | Data sharing policies | www.bbsrc.ac.uk/organisation/policies/position/policy/data-sharing-policy.aspx

Economic and Social Research Council | principles and purpose | www.esrc.ac.uk/about-esrc/information/data-policy.aspx

Engineering and Physical Sciences Research Council (EPSRC) | policy | www.epsrc.ac.uk/about/standards/researchdata/

Natural Environment Research Council | policy and principles | www.nerc.ac.uk/research/sites/data/policy/data-policy.pdf

Research Councils UK | principles | www.rcuk.ac.uk/research/datapolicy/

Science and Technology Facilities Council UK | policy | www.stfc.ac.uk/resources/pdf/stfc_scientific_data_policy.pdf

UK Open Research Data Forum | policy and forum | www.royalsociety.org/~media/Royal_Society_Content/z_events/2014/2014-01-21-Forum-Note.pdf

Furthermore, more information can be found about other initiative in other countries and open data projects.

Open data advocates

Examples of open access to research data activities in other countries:

Austria | Australian National Data Service | www.ands.org.au/

Sweden | Swedish National Data Service | www.ands.org.au/

United States | VS government | www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Activities of open data at the European / international level:

DataCite is an international organisation which aims to (1) establish easier access to research data, (2) increase acceptance of research data as legitimate contributions in the scholarly record, and (3) to support data archiving to permit results to be verified and re-purposed for future study. | www.datacite.org

LERU (League of European Research Universities) made a statement on / overview of open access to research data | www.leru.org/files/general/Open%20Access%20to%20Research%20Data-FINALdocx.pdf

PLOS (Public Library of Science) is “a non-profit publisher and advocacy organization founded to accelerate progress in science and medicine by leading a transformation in research communication.” | www.plos.org

RECODE project is a European FP7 project and the policy project “will leverage existing networks, communities and projects to address challenges within the open access and data dissemination and preservation sector and produce policy recommendations for open access to research data based on existing good practice.” | <http://recodeproject.eu/>

Science Metrix made a study about ‘open data access policies and strategies in the European Research Area and beyond’. | www.science-metrix.com/pdf/SM_EC_OA_Data.pdf

Summary | Most other countries do not have much experience with open access to research data. The United Kingdom and the Netherlands are frontrunners in the process of opening up research data. Especially the United Kingdom has several initiatives as a forum and data curation centre. At supranational level, the European Union made agreements in its Horizon 2020 programme on what data should be available and opened a pilot in order to gain experience about the implementation of open access to research data.

In each section a summary has been presented, therefore there will be no final summary of this chapter. The chapter presented a wide overview of the open access to research data and in the following chapter the theoretical approaches of chapter 2 will be applied to the input from this chapter. The following theoretical chapter, part two, will present a new theoretical framework of open access to research data and regulation.

5. Theory | part two

This chapter offers a theoretical analysis of the empirical data from the Delphi-analysis, interviews and documents. This chapter can be seen as an extensive theoretical feedback and does not aim to propose a full new theory. However, the chapter focuses on the governmental role as proposed in the first theoretical part and is based on the answers and input from the Delphi-analysis, interviews and documents. This chapter compares empirical data from chapter 4 and theory about policy instruments from chapter 2. The chapter starts with underlying arguments for intervention in open access to research data (5.1). Information about instruments and tools for establishing open access to research data will be combined with the theory of Hood and Margetts (2007) on detectors and effectors (5.2). Additionally, tools will also be related to the theory on sticks, carrot and sermon (5.3). Governmental intervention strategies will be applied to information from the analysis based on four strategies of Balch (1980) (5.4). The following sections discuss hybrid regulation of open access to research data (5.5) and open data regulation in digital age (5.6). This chapter concludes with a summary (5.7).

5.1 *Arguments for intervention*

Experts of the Delphi-analysis and interviewees mentioned several instruments and tools for regulating open access to research data. Bemelmans-Videc (1998) made a theory that explains the underlying arguments or values for intervention. The choice can be based on effectiveness, efficiency, legality and democracy values (Bemelmans-Videc, 1998). The experts and interviewees conclude that there should be intervention at several different levels and in parallel. Additionally, some kind of regulation is necessary (without specifying regulation at this moment, see more information in the following sections), according to several experts in the Delphi-analysis. The growing demand for open access to information and issues, such as privacy of individuals and the growing amount of data, calls for coordination or regulation (Guibault & Wiebe, 2013). Also the EC calls for coordination and common open access definitions in member states (European Commission, 2013a; 2013b). The goal of regulation should be to realise accessibility to research data. This can be seen as an effectiveness value, because it is about the extent that objectives have been achieved (Bemelmans-Videc, 1998). Furthermore, policy instruments may speed up the adaption of open access to research data, which increases the effectiveness. Additionally, facilities for researchers to manage data in a proper way is the overall aim of any intervention. This can be done by the government or other parties, according to the experts and interviewees. These aims reflect the effectiveness value, because its main goal is to achieve the objective.

Particularly the interviewees (Kolman, interview; Van Loenen, interview) argue that governmental regulation is not preferred if stakeholders can arrange the open access to research data together. Only when the stakeholders cannot deliver proper accessibility to research data and other related services like protection of rights or databases, should a government interfere. The government should anticipate on the current 'open movement', and should consider what to do. The question of what the government should do is therefore appropriate, according to the experts and some interviewees. This approach is more legitimate than intervention by default, because the governments does not fully get involved in the private/commercial domain (Kolman, interview; Van Loenen, interview). This argument may fall under the 'democracy argument' of Bemelmans-Videc (1998), because it is about legitimacy of policy instruments and the extent governmental intervention will get accepted by stakeholders. The general opinion in the Netherlands is that the government only should intervene if necessary (Van Loenen, interview) and should not interfere businesses and industries in their aim for innovation. The aim for effectiveness and the general thought of 'governmental invention only if necessary' form the basic values for the policy instruments for open access to research data. In addition, the value of legality has been mentioned, because formal rules (like intellectual property rights) are ambiguous and not sufficient for dealing with open access to research data. This 'legality value' is therefore important in the process of open access to research data and could be an underlying argument of possible governmental intervention. The efficiency value has not been mentioned explicitly by the experts and interviewees as the central goal or aim of open access to research data and governmental intervention.

In general, the choice for intervention (via policy instruments) in this study is mainly based on effectiveness and democracy arguments, which is based on information from experts and interviewees (Bemelmans-Videc, 1998).

5.2 *Detectors and effectors*

The approach of Hood and Margetts (2007) distinguishes two purposes of policy instruments: detection (governmental tools for taking information) and effectors (governmental tools to influence society). The analysis highlights several governmental instruments in the sense of effectors (Hood & Margetts, 2007, p. 3). Firstly, the government should take up its role as **promoter** of open access to research data. It should respond on the open access movement and make to some extent an impact on society. Also detectors purposes can be found in the analysis. The government should get **informed** about open access to research data, so partly it may use tools like detectors in order to understand what is going on in society related to open access to research data (Hood & Margetts, 2007). This is reflected by the proposal that a government may create an environment where people (researchers, librarians and repository/data managers) can be trained properly and foremost bringing

different researchers and disciplines together. These kind of interventions are detectors (receiving input via training and meetings) and effectors (reflecting a governmental approach via training and meeting to influence individuals and organisations). Furthermore, assistance in providing access to data for a more general population by **informing** them is also a role for the government, according to the experts in the Delphi-analysis. It is about affecting the society and to being able to perform these roles a government should be informed itself. This could be done by organising meetings and taking part in discussions and further development of institutions that work with open access to research data.

In particular, the effectors role is reflected in the coordinating role of a government. As said before, the detectors and effectors role do overlap each other because coordinating and informing can only be done if the government is informed. One of the roles that influences society is that government should ensure that stakeholders cannot create a monopoly on open access to research data. Moreover, the government should take up its responsibility to have a more regulatory role in disciplines or sectors where the supply of open (research) data is lacking. If there is no demand for open access to research data, governmental intervention is not necessary. Examples of disciplines that do not demand open access are unknown. An example of a discipline with a high demand for open access to research data is the geo-sector (mentioned by one expert).

As mentioned before, another governmental role is to take **legislative measures**. It is about the necessity to tackle legal obstacles and uncertainties through legislative measures. The intellectual property rights and database rights do not fully cover the rights of the stakeholders in open access to research data. In order to have an effective policy, it is crucial to be very specific with regard to a patenting-compatible schedule in mandatory clauses. This also applies to the contribution percentage and company profiles. In addition, “adopting open access calls for improvement and acceleration of the classical investment exploitation directive.” The current situation is that once the data are in the public domain, no royalties can be obtained for them nor can patents be obtained and therefore exclusivity of exploitation rights cannot be granted. Speeding up the process of intellectual property through an improved decision process can be beneficial to valorisation through classical investment exploitation. This study did not go deeply into this topic and did not highlight all the barriers in detail. More research should be done in order to have a full understanding of this effectors role by the government. The government should establish a framework for open access to research data that may function as a guide for working with open access to research data. To summarize this theoretical thought on governmental regulation is that a detectors role should be taken up and in parallel an effectors role by coordination the process of open access to research data.

5.3 *Sticks, carrots and sermons*

The theoretical approach of sticks, carrots and sermons highlights some different aspects of governmental tools in open data (Hague & Harrop, 2010; Vendung, 1998). The policy instruments can be classified in three types: regulation (sticks), economic means (carrots) and information (sermons). This approach can be used to categorize the regulation strategies of the (Dutch) government.

One of the experts explicitly mentioned this approach for explaining his perspective on governmental intervention (without knowing the theoretical background of this study). The expert proposed a mixture of carrots and sticks by dividing it into a funder and **legislator** role. This perspective and other insight of experts and interviewees are combined to make the following regulation theory on open access to research data. First of all, sticks (regulation) are used in the governmental legislator role. The aim is that all relevant stakeholders (for example, researchers, research organisations, universities, funders and publishers) should act according to certain legal requirements. The legal requirements could entail (mentioned in the analysis): (1) a requirement that all research data produced within public research institutions has to be made accessible in an appropriate way, irrespective of the research funder, (2) a requirement that every publically funded research institution has to have a certified data repository that fulfils certain interoperability standards and is linked to a national data centre, et cetera, or (3) that every recognized programme for a Research Master or a PhD has to include a module on research data management.

Furthermore, in this context the stick is non-eligibility of costs if certain requirements are not fulfilled. This can be part of the regulatory framework in case of non-compliance of researchers. This stick also aims to convince the private sector and industry to comply to open access to research data. Stakeholders who do not comply will not be able to benefit from public funding or subsidies if they follow the basic principles of openness. A maximum deadline for making data accessible (embargo period) can be established/demanded and if researchers or research organisations. If there is no compliance, sanctions can be implemented (for example, non-eligibility of costs).

Another stick is that the government should provide a supportive **legal framework** and should establish a general **regulatory framework** for dealing with open data (based on the Delphi-analysis, section 4.9). The framework should be flexible and give guidance instead of strict enforcements. Additionally, the government sets specific policy standards using soft law measures and in some cases mandates. Non-enforced mandates are very ineffective tools and therefore monitoring is important. For example, many funders have policies regarding open access to publications and increasingly also regarding research data, but there is a low level of compliance because sanctions are rarely imposed. This also includes the prevention of a monopoly on open access to research data and taking up the responsibility to have a more regulatory role in disciplines or sectors where the supply of open (research) data is lacking. Furthermore, the government should

take into account the interests of businesses and industry, just like the European Union does in its open access to research data pilot. A recommendation is that the establishment of a regulatory (legal) framework should be done incremental, which will allow more flexibility and increases the success of open access to research data. Start for example with one domain and then expand.

An important regulatory intervention is the acknowledgement of publishing research data within the career system. However, this is not specifically a governmental role, but merely a role for the academic/research sector. But the government can encourage this.

Secondly, the carrots (economic means) are reflected in the governmental role as **funder** of open access to research data. Additional funds, grants or subsidies have been emphasised by several experts and interviewees. Financial support for researchers or research organisations may help to finance the investments for data infrastructures. It could be an incentive to deposit and share data, because current data cost are high. Furthermore, 'carrots' could be additional support for activities related to data depositing, maintenance and storage, including taking security measures. Visibility can also be increased through additional funds. In particular, financial incentives can be beneficial for data that need to be stored for long-term. For example, data that should still be accessible within fifty years could be very costly. The data infrastructure should be updated according to new soft- and hardware and this is very costly. Additionally, data storage for a long period and its maintenance is also very costly. The government can support these infrastructures and maintenance via financial incentives. However, data maintenance and infrastructures can also be done by co-financing. The research institute concerned with gathering the data and the government can both finance the data system to make it accessible to the public. Facilitation is preferred because subsidies may create time for building a dataset by stakeholders. In most cases the finance of research ends when a study finishes. By giving subsidies, researchers can work longer on proper data recording and its storage. This is often not part of the research finance, because deadlines are often too strict.

Given the current economic climate a strong political commitment to support is required to the allocation of funds for data related issues, while funding is cut in other areas that may seem more relevant and of more immediate benefit to some stakeholders. Disincentives like taxes or fees have not been proposed by the experts and interviewees.

Thirdly, sermons (information) are reflected in the **coordinating** role of the government. The government should create awareness about open access to research data and issues as DMPs and the possibility of a credit system for researchers if they publish their research data. The creation of an environment where stakeholders (researchers, librarians and repository/data managers) can get information about establishing a DMP and to prepare their data for an easy process of publishing research data online (by using certain formats for an easy and proper data recording which can be uploaded without many alterations). Creating awareness can be established via training or (network)

meetings and seminars. Researchers may also hesitate to share research data, therefore best practices can be used to convince researchers to share their data. Researchers and other stakeholders need to see and recognise the benefits of open data. Furthermore, consensus and common ground can also be established via training and meetings. It is very important that all stakeholders know what each of them is doing or can be asked to do and they may join forces to tackle problems or act together in stimulating open access to research data. The government could play an important role in facilitating these kind of meetings and training by bringing together different researchers and disciplines. The 'helicopter view' of the government can be beneficial in establishing common ground or maybe even more or less governmental regulation, because government can set up a strategy for establishing open access to research data for the long term and for all stakeholders. Coordinating meetings and training is an important governmental role. A specific aspect during open access to research data meetings is giving information of providing access to research data via advice of which infrastructure could be used and which IT-solution works best.

In addition, the government should stay informed, itself as well because unexpected regulatory factors may play a negative role in the success of open access initiatives, such as contrasting IP rights, like ownership. Definitions and legislation should be consistent in order to establish open access to research data. In general, the government should take up a role as a watchdog and responsiveness instead of a role as director of open access to research data.

The feasibility of the regulatory framework depends on the willingness and commitment of all stakeholders, including a monetary commitment for financial support. It will also be crucial to have a mechanism for a continuous open dialogue between the government and the different stakeholders concerned. In this way continuous feedback of the developments in open access to research data can be discussed and if necessary action can be taken.

5.3.1 *Other regulatory levels*

Several sticks, carrots and sermons can be recognised in the process of open access to research data, but not all are not specifically a governmental role. Sticks can be recognised at the level of funders as well. They should also set policy standards (nationally and internationally) using soft law measures and in some cases mandates. One of the standards can be the demand for open access to research data as a requirement for funding (like the EC does), similar as the governmental stick of no financial help in case of no open access. Furthermore, at the European level rules could be necessary to establish open access to research data. The pilot on open access to research data is an example of the path to possible regulation and is therefore an appropriate way of exploring the effects of possible regulation, according to several experts and interviewees.

Carrots should mainly come from governmental level, according to the experts and participants. However, as said before, funders can demand open access to research data as a requirement for funding. This is a disincentive and could enforce researchers to give access to their research data.

Sermons can be used at the institutional level. Institutions such as research organisations and universities need to invest in support for researcher. In their technical skills and the amount of employees. This entails standards for citation, advocacy, expert advice on legal issues, technical information and metadata. Furthermore, advice could also focus on the DMP and how to store data during and after a study. Institutions can also provide researchers conversion tools to open formats, which eases the data storage and recording. Additionally, tools for researchers to choose the correct licence for their data can also be provided by institutions. The institution can provide workflows and processes to include content in their own repository or a national / subject-specific one. In general, institutions should create trust and legitimacy of research institutions and research findings among researchers. They should present data in the right way in order to gain support of researchers. Anonymising data can increase trust among researchers. The research team should anonymise data from the beginning of a study. The institution can also take up the (meta)data quality control, but also the data repositories can take up this role. More details on the implementation of this quality control is missing.

At the level of funders and publishers, sermons can also be recognised. A coordination role can also be assigned to funders and publishers, according to experts and interviewees in the Delphi-analysis. Funders and publishers can also initiate meetings and training about open access to research data for researchers and other stakeholders. During these meetings a common drive needs to be set up by all stakeholders who have a vested interest in creating and maintaining the dataset. This because of several contrasting perspectives and definitions. For example, some stakeholders want open access to specific and detailed information, others do not want to give their detailed data because of their competitive edge. Furthermore, similar definitions are necessary to make progress in the process of open access to research data, otherwise contrasting definitions can hinder the establishment of a regulatory framework. Moreover, funders and in particular, publishers can take up a coordinating role in giving access to the data underlying publications, just like journals. Also funders can take up that role by communicating that non-compliance of open access to research data means non-publishing. Furthermore, publishers can help researchers to deal with research data, in particular by helping to describe (meta)data. Digitalisation of data and data recording is also familiar to publishers and they could inform researchers also on this issues.

5.4 *Strategies*

As mentioned in the first theoretical part of this study, four strategies for imposing policy instruments are possible (Balch, 1980): informing, facilitation, giving incentives and regulation. The first strategy is one of the most explicit mentioned strategy or role for a government, which is already mentioned in the sections before. The government should inform stakeholders in the process of open access to research data. Experts and some interviewees emphasise the coordinating role of a government and the establishment of an environment of interaction, trust and reliability. Establishing a framework, as specified in section 4.9, can also be recognised as an informing strategy, because it may give the boundaries, discretion and focus of open access to research data. Clarity in open access to research data is needed, especially for businesses and industry. Several questions are still not formally answered by the experts and interviewees, for example:

- Which data should be openly accessible?
- Who is the owner of data?
- What is research data?
- Who decides on which data are relevant to maintain?
- Who will pay for data maintenance?

Therefore, the government could take up a role by **defining concepts and definitions** in order to establish clarity. However, interaction by receiving input from stakeholders is essential for establishing a dense and supportive framework. The government should highlight the ‘added value’ to a product by focusing on opportunities and moving further with the open movement of this era. Information motivates change (Balch, 1980) and researchers often need to be convinced, because they have several arguments for not opening up their research data (see section 4.7). Information and a framework should also be discipline and field dependent, with less uniform guidelines and applicable to several circumstances.

Secondly, facilitation strategies are also mentioned by the experts and interviewees. This closely relates to the aforementioned strategy, because informing stakeholders may result in convincing them and changing their behaviour. The most essential facilitation strategy is the **funding** aspect in open access to research data. Difficulties should be removed and especially high costs related to data storage, maintenance and data infrastructure in general should be decreased. According to most of the interviewees and experts of the Delphi-analysis this should be done by the government. Additionally, changeover cost or specific required facilities can be overcome by governmental assistance (Balch, 1980).

Thirdly, incentive strategies can be recognised in measures like reducing the costs and in general by providing a framework for establishing and dealing with open access to research data. Criteria at governmental level (mentioned in section 4.8.1) were: establishing a suitable framework

for open access to research data, legislative measures, financial support and policies to stimulate open access to research data. Furthermore, create an environment where people can be trained properly, bringing together different researchers and disciplines, and ensure that stakeholders cannot create a monopoly on open access to research data. In general, these aforementioned criteria do function as a reduction of the costs of obtaining information and adoption of facilitation strategies more easily. The grants and recognition of publishing research data will also function as incentives for researchers. However, the latter is not a specific role for the government.

Fourthly, regulation strategies should be **modest**, according to experts and interviewees. Government should take up an legislator role, but should not intervene if there is no need to. This regulation strategy overlaps the regulation approach (sticks) of Vendung (1998). However, Balch (1980) emphasises the legal nature of the regulation strategy, including punishment and monitoring of behaviour. Constraining and ambiguous legislation, like ownership, should be more clear. Furthermore, agreement on definitions by stakeholders is necessary in order to establish open access to research data. Sanctions are proposed by not providing funding if stakeholders do not comply to the principle of open access to research data. So, compliance and sanctions are proposed to establish open access to research data. Moreover, an agency needs to do the regulation in order to enforce compliance (Balch, 1980). Experts recognise the government as a watchdog in the process of open data.

In general, the role of the government should not be intervention by default in the context of open access to research data, rather a more balanced role. If stakeholders can arrange agreements and further implementation themselves, then the government should not intervene in these circumstances. The governmental role should be a more of a watchdog and respond role instead of a director's role. Furthermore, funders should regulate the open data themselves by demanding open access as a requirement for funding. This is related to the carrot and sermon approach. Furthermore, information and facilitation work best on 'motivated' people (Balch, 1980). Therefore, the government can set up an environment to create trust and to motivate stakeholders. Monitoring is essential for a reliable and efficient strategy on open access to research data.

5.5 *Hybrid regulation*

As stated before, legitimacy in open access to research data and the policy instruments is highly important in the actions that governments take. Using hybrid regulation can increase the legitimacy of regulation of open access to research data. Firstly, **co-regulation** is necessary, in the sense that all stakeholders should take up a role in the implementation of open access to research data. As mentioned before, institutions (like research organisations and universities), funders and publishers need to participate in the process of open access to research data. These stakeholders can invest in

technical infrastructures and staff. This entails advice on – for example – standards for citation, DMP, legal and technical issues, data storage and metadata. Another example of co-regulation is a possible reform of the career system of researchers. Research data should be acknowledged as publications and/or open access to research data should be a requirement in evaluation criteria for research projects and PhD admissions. Additionally, each PhD programme should entail a module on research data management. Moreover, funders and publishers can also have a coordination role in open access to research data. These stakeholders can also initiate information meetings or training for researchers. Together the stakeholders exchange information on open access to research data and consensus could be established.

Secondly, **self-regulation** can be recognised in the way that funders already make their own policies on open access to research data. For example, the demand for open access to research data as a requirement for funding for researchers. Additionally, soft laws from the government could facilitate funders in making their own policies on open research data and implementing them. The government encourages open access to research data in this way by giving a regulatory framework, soft law and flexible regulation. Publishers and other stakeholders can then implement these aspects according to their field or discipline in order to have an targeted and appropriate approach of open access to research data.

Thirdly, **meta-regulation** is an abstract term in the process of open access to research data. Meta-regulation can be identified by implementing information meetings or seminars. This kind of regulation (information) will verify or test the soft laws or regulatory framework in practice. Input during these kind of meetings will give information about the regulation and its practicability. It could give feedback to the government about its own regulation strategies, 'it regulates the governmental regulation'. Furthermore, the facilitation strategy will also give information about the workability of the regulatory framework and other soft laws. For example, the degree of use of facilities can be an indicator for that.

Fourthly, **multi-level regulation** is desirable in the process of open access to research data. In particular, supranational (soft) regulation would be appropriate, because this may increase the reciprocity of sharing data. Furthermore, research is increasingly more international oriented. The European Union can be seen as a player against other continents and can therefore maybe enforce or encourage data sharing of other continents and its research organisations. The EC pilot on open access to research data is a starting point of a supranational regulatory framework. This can be seen as the first steps in open access to research data in the European Union, according to one interviewee. At the supranational level the European Union should cooperate with national and sub national governments. Finding consensus and an appropriate approach is beneficial for the implementation of open access to research data. Furthermore, the outcomes of the EC pilot are very

important to learn from experiences on open access to research data in European projects. Based on these experiences future regulatory steps can be taken. Furthermore, national regulation is important by drawing a regulatory framework and other facilitation and information strategies, as mentioned before. The government can take up different roles, such as a funders and informing role. At the institutional level, national and local repositories play a role in data storage and maintenance. Organisations that work on data can be nationally or locally oriented. These organisations are essential in the implementation of open access to research data. In particular, researchers should choose the best option for depositing their data, this can be for example at local or national level. Therefore these institutions are essential in the process of open access to research data by providing infrastructures and/or advice.

These aforementioned hybrid regulation approaches could be more effective than imposing rules, because open access to research data can be established by encouragement and trust in the stakeholders. Therefore these hybrid regulations are appropriate for open access to research data.

5.6 *Digital era*

Based on the analysis, the focus of a government in the digital era should be on using its **nodality** position (Hood & Margetts, 2007). This has been recognized by experts and several interviewees by mentioning an ongoing interaction between stakeholders of which a government can take up a leading role (the information strategy or sermons). The government may function as coordinator in the discussion and further implementation of open access to research data. As mentioned before, the government should be in the middle of the social network (nodality) in order to use its role as disseminator of information for changing the behaviour of stakeholders in research data (Dunleavy & Margetts, 2000; Hood & Margetts, 2007). The government should be active by being visible and taking part of meetings on open access to research data. By initiating meetings, seminars and training, the government can fully take up its nodality role. The governmental perspective can be disseminated via these meetings. In particular, the **open dialogue** can be set up in order gain legitimacy for governmental regulation, like a possible regulatory framework. Additionally, by initiating and participating in meeting on open access to research data, the government receives information from the stakeholders and different sectors. This information could be useful in formulating policies and (soft) laws or drawing additions to existing policies. Again, legitimacy of governmental regulation could rise if all stakeholders are involved by giving input and best or worse practices.

As mentioned in the first theoretical part, governments could ensure that its own website on open access to research data are listed in the top ten or twenty when a search query is made through a search machine. Clarity is also highly important, as shown by the analysis, experts and interviewees

state that there is a lack of common ground and clear definitions of open access to research data and other related topics, like research data and ownership. So, clear definitions can be defined by working together with all stakeholders via – for example – meetings and seminars. These definitions can be disseminated as a governmental communication and published on governmental websites. This kind of information should be easy to discover on the website in order to disseminate its perspective as much as possible.

The government should be able to use its instruments and communication methods at a variable range of intensity (Hood & Margetts, 2007). The use of instruments should be flexible and targeted. This can be done by using – for example – different levels of financial support and targeted at researchers. Further details of this cannot be given based on this study. Additionally, policy instruments need to be direct (Hood & Margetts, 2007). Examples of policy instruments that are direct: financial support for researchers (sermons) or the sanction of non-eligibility of costs by non-compliance of open access to research data among researchers. Summarising the aforementioned, regulation of a government in a digital era needs to be **active, informative, targeted and flexible**.

5.7 *Summary*

Open access to research data can be approached from different theoretical perspectives. There should be intervention at several different levels and in parallel. Governmental intervention can only be done if necessary. If stakeholders can arrange the process of open access to research data, then governmental intervention is not desirable. Arguments for interventions are based on efficiency, democracy, and legality.

The government can take up a detectors role (staying informed) and an effectors role (affecting behaviour by initiating meetings on open access to research data). Governmental regulation strategies of open access to research data are sticks (regulation) by taking up a legislators role and by creating a regulatory framework. Non-eligibility of costs can be sanctioned if there is non-compliance in the case of open access to research data. Carrots (economic means) entail financial support for researchers. Furthermore, sermons can be recognised in the coordinating role of a government in the open access to research data process. Governments can inform stakeholders by initiating meetings and training and where consensus/clarity can be created on several issues like what data should be open and who will pay for maintenance of data systems. Governmental facilitation strategies are also desirable, by funding data storage, maintenance and/or data infrastructures.

Regulation of open data can also be done at other levels (by institutions, publishers and funders). These kind of hybrid regulation strategies can be: co-regulation (government and stakeholders), self-regulation (stakeholders, like funders establish own policies), meta-regulation

(meetings can regulate the governmental regulation by giving feedback) and multi-level regulation (at European level and using local or national (private) institutions). The open access to research data is part of the increasing digital era and governments should therefore take up its nodality position. This entails that the government should be in the middle of society and networks of open access to research data. The government should be visible and initiating activities on open access to research data.

The following chapter presents the overall conclusion of this study.

6. Conclusions and recommendations

This final chapter answers the general question of this study and gives an overview of the circumstances of the open access to research data process. This study foremost maps the contemporary situation of open access to research data and its different perspectives. Firstly, the sub questions will be answered and secondly, the general question will be answered (6.1). This section concludes with recommendations based on the results and my own insights. Followed by a substantive and personal reflection (6.2) and recommendations for future research (6.3).

6.1 *Research questions*

The central research question was defined as:

To what extent are possibilities available and desirable for the Dutch government to regulate open access to research data?

Based on this central research question, open access to research can be decomposed into several research questions, which is listed as follows:

1. *What is open access to research data?*

There are several definitions of open access to research data. First of all, research data can be seen as an ideal, which entails raw data, methodology data, process data, processed data and associated metadata. Currently, there is no consensus on what research data are. More clarity on what research data is would be desirable. In addition, the definition of research data depends on the discipline of research. Basically, it is about the facts that need to be interpreted (like raw data, processed data, interpreted data, data coming out of a machine, a series of photos, of documents, and questionnaires) and data that have been collected or generated during the research process (through – for example – experiments, observations, measurements, simulations, calculations and surveys). Definitions of methodology data, process data and processed data should be defined sharper. The definition of metadata can be data that provide information on key features of the research data. A clear definition of raw data is currently missing.

Secondly, open access is the right to access and re-use something. Related to research data it is the right to access and re-use digital research data. Data can be accessed, mined, exploited, reproduced and disseminated free of charge for the user (European Commission, 2012b). The terms ‘open by default’, usability and reusability are important in the debate on open access to research data, according to the experts. Furthermore, data from negative results should also be accessible. Access can be given to raw data, data underlying peer-reviewing publications, data which are hard or

impossible to generate or extremely expensive (based on the Delphi-analysis). Moreover, access could be given to metadata and all data that is necessary to interpret and re-use data and which enable verification.

Thirdly, definitions on open access to research data are also related to issues like sensitive data and commercial data. Defining sensitive data is – again – a difficult issues. The domain of sensitive data is getting more difficult to define, because of larger data sets and the combination of data sets. Sensitive data could apply to privacy, confidentiality and security issues. Furthermore, the definition of commercial data could be seen as a sensitive topic. In general, commercial data are all data that businesses and industry make a profit on.

Finally, open access to research data is often related to publicly funded research. In general, this kind of research is all research which is (partly) funded by the government. Some people state that research – which is funded by the government – should be accessible to all. However, businesses and industry cannot agree on this. Perhaps a definition of public research data in the context of open access should be formulated to establish more clarity in the private sector.

2. *Is there a need for open access to research data?*

Open access to research data is part of a more general ‘open movement’. The requirement for open access increased through the growth (wealth) of data in governments, research institutes and (other) public bodies and organisations. Furthermore, digitalisation and the dependency of information and communication technologies increased the requirement for open access to research data.

The general opinion is that research data need to be accessible for the public, for each individual. It is about “democratisation of data”. Especially research (partly) financed by public means should be available to all, with the exception of sensitive data. In particular, many citizens feel that these data should be freely accessible; it is partly a problem of ‘entitlement’: people feel they have the right to access the data. Furthermore, societal and economic issues are tackled not only by public bodies, but also by private actors. Granting access to (research) data would further help in enabling these actors to not only make money by building applications or a business on these data, but also solve societal problems.

Research data is probably foremost important for the research community. Ordinary people perhaps do not understand or are not interested in the data. Probably only a few want to have and need access. Additionally, data are not self-explanatory and are a matter of context. A suggestion is that opening-up the data to everyone is easier and more cost effective than a complex system which gives access to specific actors, but a modest approach is also suggested. Open access to research data could lead to better and more relevant research output, but such output has many variables so

it would difficult to legitimise open access to research data with that. In general, if there is no access to research data this will not make such a big difference for the general public.

3. *What are confronting opportunities and difficulties in the process of open access to research data?*

Open access to research data is a new debate and should be seen as a principle and an opportunity for society. Easier access to research data can contribute to solve several problems that society is struggling with. The issue of poor efficiency of research data can be tackled by re-using data and open access can increase the quality and efficiency of scientific research. Furthermore, open access facilitates research integrity. In particular, validation, verification and evaluation of research (results) will be facilitated by publishing underlying data. Moreover, legitimacy of and trust in research organisations and research finding can be created because of open data. Open access to research data can also be beneficial by boosting data sharing and re-use of data. The possibility to use data and to combine datasets creates larger datasets that are in principle statistically more robust. Another advantage of sharing and combining different datasets is the possibility to answer other or more (research) questions than would be possible to answer with the separate dataset only. Additionally, accessibility to research data and the possibility to combine datasets could support innovation for businesses and industries.

Besides the opportunities, open access to research data does have (potential) down-sides. A barrier can start with researchers that created the data who want to hold the data on to them for reasons like 'it is my data that I worked very hard to collect' or 'I cannot trust the data produced somewhere else'. Furthermore, people could fear that 'everyone' is being able to see what goes on in one's public or private organisation. Issues related to privacy, confidentiality or security could form a barrier in the process of open access to research data. Furthermore, there is a lack of awareness for the benefit that open research data could bring to the science system. Not all researchers are familiar with accessibility to research data and its potential. Sometimes legal obstacles and uncertainties – for example in data mining – form barriers for open access to research data. The infrastructure of data systems is a frequently mentioned problem in open access. Stakeholders, like researchers are not always familiar with the technical aspects of data depositing, storage and security. Providing access and publishing of data can even form problems for researchers.

Moreover, the costs of publishing data and the maintenance of data after the data are published can be constraining. Several of the aforementioned problems are rooted in the lack of training in open access to research data, as well the lack of clear standards, licensing and protocols in open data. Furthermore, there is no consensus on several definitions related to open access to research data. One of the ambiguous definitions is the definition of ownership, which has been

discussed extensively in the analysis. However, the discussion who is the owner of data is ambiguous and probably not a very productive discussion, because opinions depend on the perspective taken. For example, ownership in legal and moral perspectives differ in thinking in terms of the main processor. Questions arise like: who needs to give access and who can decide on that? The general opinion is that access to research data should be given by the institute and the researcher. However, co-ownership and contracts often define the authority for giving access to research data. In order to clarify ownership, the data management plan should entail information about the owner of data and its rights. These kind of ambiguous definitions complicate the debate and process of open access to research data.

Additionally, lacking incentives, recognition and harmonisation may increase the barrier for publishing and sharing research data. Unfamiliarity with open access to research data constrains the process for more accessibility. Unfamiliarity can be decreased by exchanging information in a community, since sharing data could create new (research) communities or groups. However, the implementation of these communities is lacking. Another difficulty of open research data is that data cannot be easily verified, tested or deeply understood without context and provenance information of the data. Questions about open access to research data still exist, like: what is the certain data set?, and who worked on it? In general, difficulties consists of unfamiliarity, lack of training and the costs of publishing and maintaining data.

4. *To what extent is there a need for a regulatory framework on open access to research data?*

In general, there is agreement on the necessity of intervention and a regulatory framework on open access to research data. The main focus of the framework needs to be, as open as possible, closed if needed. The default should be openness, but sometimes it is not possible because data are sensitive. The term 'intelligently open' could therefore be appropriate in this context. According to experts, data that should be freely accessible are: data underlying publications, data needed for validating results, metadata, data which have been specified in an agreement and information about tools at the disposal of the beneficiaries. The framework can be based on the approach of the European Commission in its open access data pilot in Horizon 2020. This approach has been regarded as appropriate to experts and several interviewees.

A final regulatory framework should be discipline dependent. A 'one size fits all' cannot be used, because regulation fits not all different disciplines. For example, biomedical research has little in common with data collected in the context of research on medieval history. Furthermore, a framework should harmonise contradictions in open access between funders and the definitions in use, because such a variety in definitions creates complexity for all stakeholders. For example, ownership should be defined or at least a manner of which ownership should or can be defined in

each research. Declaring an owner of data in the data management plan as a requirement could be suggestion for improvement. Additionally, mandatory clauses according to intellectual property dissemination and exploitation in EU-funding contracts should be maintained. For example, ownership is part of this intellectual property legislation and mandatory clauses should be established in order to legally cover this.

An essential aspect is a data management plan, additional with specification of a licensing scheme (like Creative Commons) and recommendations for trusted repositories. More specification of what is involved is necessary in order to meet appropriate open access to research data. Clarity on data authenticity is necessary; who need to control and verify data before publishing? Agreement does not exist on this issue. Another issue of the framework should be that data are stored properly. This can be established via national, institutional or discipline dependent (trusted) repositories. Discoverability, protection of data and accessibility for the long-term should be the principle. In addition, data management do requires a lot of funding, in particular if data should be kept for a long term. Financial support is therefore necessary. The extent of maintenance depends on the kind of repository or data storage, but there is no consensus on who need to maintain a dataset and/or who will pay for these additional costs. Currently, data maintenance is often not part of the research budget and therefore data maintenance and accessibility for the long-term cannot be ensured.

Besides, there should be no extra costs for accessing research data and the framework should not have strict formalizations. Businesses and industry also stress the latter. Furthermore, commercial interest of businesses should be taken into account. Businesses and industry ask for an opt-out option and want to decide for each research project on a case-by-case basis whether or not the research consortium can give open access to its research data. Opening all data to everyone will be problematic for business and industry.

Moreover, a decision should be made on the necessity of an embargo period after which the data becomes openly accessible. Some experts emphasise the advantages of it, other experts disagree on that. In general, the framework should create an environment of awareness, trust, recognition and support for researchers. As mentioned before, researchers often lack the trust in research organisations and result. Therefore a regulatory framework should establish compliance procedures and recommendations on trustworthy digital repositories. Furthermore, agreement on definitions is again essential in creating trust. Reciprocity should be ensured, but this is a complex issue to enforce. Institutional recognition is also important to create trust among researchers. Several stakeholders as the government or institutions (universities or research organisations) can encourage this change of culture in science. Another essential part of open access to research data is that researchers should be trained properly. Awareness should be created and knowledge on data recording, storage and maintenance should be established. This could increase the trust and use of

open access to research data. if researchers do not comply with the requirements for open access to research data, then sanctions can be taken. As mentioned before, non-eligibility of costs can be an example of a sanction.

The feasibility of this framework depends on willingness and commitment of all stakeholders and it will be crucial to have a mechanism for continuous open dialogue between the government and the different stakeholders concerned. This emphasises the coordinating role of the government, but also the initiating and participative role of other stakeholders. An incremental approach, probably even starting with giving access to one domain, is suggested for this complex process of open access to research data. The implementation of open access to research data will take time.

5. *What are possible types of regulation of open access to research data?*

Regulation of open access to research data can be approached by different theoretical perspectives. In general, there should be regulation at several different levels and in parallel. Governmental intervention should only be done if necessary. If all stakeholders can arrange and implement the process of open access to research data, then governmental intervention is not desirable. In case of intervention argument are based on efficiency, democracy and legality reasons. The goal of regulation should be on realising accessibility to research data, this can be seen as the effectiveness value. Policy instruments could speed up the adaption of open access to research data. Additionally, facilities for researchers to manage his data in a proper way is the overall aim of any intervention. The notion that governmental intervention only should take place if necessary, can be labelled as a democracy argument. It legitimises possible intervention of the government. The value of legality is applicable, because formal rules (like intellectual property rights) are ambiguous and not sufficient for dealing with open access to research data. This legality argument is therefore important as justification for regulating open access to research data.

The government can take up a detectors role (staying informed) and an effectors role (affecting behaviour by initiating meetings on open access to research data). Governmental regulation strategies of open access to research data are sticks (regulation) by taking up a legislators role and by creating a regulatory framework. However, enough flexibility must be provided to take into account the specific interests of private entities. Non-eligibility of costs can be sanction if there is non-compliance in the case of open access to research data.

Carrots (economic means) or incentive strategies entail financial support for researchers. This strategy is the most frequently mentioned by experts and interviewees. In particular the funding of research data management and sharing and the associated support in an integrated way – proposals that present a well thought-through data management strategy and provisions for proper data

storage beyond the duration of the project should receive credit for this in the evaluation, and costs associated to these activities should be eligible, also for a reasonable period after the lifetime of the project if this is justified. Investments in altmetrics is recommended, however this can also be done at other levels since experts and interviewees did not specify who have to take up that role. In general, incentives are important to convince researchers to give open access to their research data. Re-use of data needs to be encouraged, for example by funders and within the academic system. One of the key requirements is to recognise data sharing as publications. A requirement for open access to research data in evaluation criteria for – for example – tenure, projects and PhD admission would also be desirable. The government could facilitate or encourage these kind of reforms in science. Other incentives for researchers are to determine common standard for citation of data. This relates to the regulation strategy by creating a regulatory framework and the facilitation of common grounds for open access to research data.

Furthermore, sermons can be recognised in the coordinating role of a government in the open access to research data process. This combination of the information and facilitation strategies could entail that the government stakeholders informs by initiating meetings and training. Moreover, these meetings can also function as platform for creating consensus or common ground on several issues, like what kind of data should be open and who will pay for the maintenance of data systems. In particular, governmental facilitation strategies are desirable in the sense of funding data storage, maintenance and/or data infrastructures.

Regulation of open data can also be done at other levels (by institutions, publishers and funders). These kind of hybrid regulation strategies can be: co-regulation (government and stakeholders), self-regulation (stakeholders, like funders could establish own policies), meta-regulation (meetings can regulate the governmental regulation by giving feedback) and multi-level regulation (at European level and using local or national (private) institutions). Open access to research data is part of the increasing digital era and governments should therefore take up its nodality position. This entails that the government should be in the middle of society and networks of open access to research data. The government should be visible and initiating activities on open access to research data. All in all, because of this coordinating and nodality position, the process of open access to research data can be regulated including flexibility.

Based on aforementioned answers, the general question of this study can be answered. The research question was defined as follows:

To what extent are possibilities available and desirable for the Dutch government to regulate open access to research data?

Open access to research data is a movement in which the Dutch government wants to be a frontrunner. Therefore, the government can use several regulation strategies, like regulation, financial, incentive and information strategies. The general opinion of participants in this study is that there is a need for some kind of intervention. However, the degree of intervention depends on the participants' perspective. In general, a regulatory framework has been regarded as necessary in order to gain clarity, reciprocity, protection, security and advice on open access to research data. The framework should be flexible and approached on a basis. A discipline dependent approach is essential for establishing open access to research data. The role of the government should not be intervention by default, rather a more balanced role is preferred by the participants in this study. The government should take up a role as a watchdog and responsiveness, instead of a role as director of open access to research data. The strategy should mainly focus on providing information for all stakeholders in order to create common ground, awareness and trust. Further (incremental) steps can be based on these principles, followed by establishing definitions and/or regulation (soft law) if possible and preferred. Many issues like unclear definitions and consequences, the lack of consensus and trust should be tackled in order to make progress in the open access to research data. Legitimacy and providing information is essential. Besides, financial incentives from the government are requested by several participants in this study. This would encourage the stakeholders to participate in the open access to research data movement. Moving together to a common goal of sharing data in order to boost research and innovation is feasible.

Based on the aforementioned information several recommendations for the Dutch government can be established in order to regulate the process of open access to research data. These recommendations for the government are as follows:

- Governmental intervention should only be done if necessary;
- Regulation should be done at several different levels and in parallel. In particular, co-regulation and multi-level governance is recommended;
- Facilitate continuous open dialogue with all stakeholders in order to establish consensus on the implementation of open access to research data;
- Initiate meetings and training for stakeholders on open access to research data in order to stay informed and to inform stakeholders;
- Provide information on aspects of open access to research data, like ownership, data recording, data storage, repositories
- Provide a flexible regulatory framework by taking into account the interest of stakeholders like business and industry. The framework can be established in cooperation with institutions, funders, publishers and researchers;

- Regulation should be on a case-by-case basis and discipline dependent;
- Provide financial support for researchers, research institutions and/or study projects;
- Require open access to research data and a data management plan in order to receive (governmental) funding;
- Strive for common ground on definitions in open access to research data. In particular definitions on research data, raw data and ownership;
- Encourage a change in the academic system in order to recognise data sharing as publications and part of the evaluation criteria.

Open dialogue and cooperation between stakeholders is essential to establish open access to research data. Without cooperation the process of open access to research data cannot improve, because stakeholders play an important role in the implementation and compliance of open access to research data. Meetings and training may lead to common ground on open access to research data. In the end this will result in successful access to research data for all and a boost for research and innovation. Cooperation is a key word in the process of open access to research data. I would like to end with a quote of Henry Ford (1863 – 1947), an American industrialist and founder of the Ford Motor Company, who expresses the path to success.

coming together is a beginning;
keeping together is progress;
working together is success.

– Henry Ford –

6.2 *Reflection*

In this section, I will reflect on some of the choices made during the study. This study has been conducted at the office of the Netherlands house for Education and Research (Neth-ER) in Brussels, Belgium. During this period at the office of Neth-ER I strengthened and improved my abilities by working at the office doing specific tasks as trainee and switching to work on this study.

One of the first things that is worthwhile to mention is that the topic of this study was very educational for me. I was not familiar with the topic and I barely knew the characteristics of open access or big data. All these kind of terms were abstract and were not part of my environment as a master student. This study has tremendously increased my knowledge about information management and research data. During the research I became more interested in the topic because of its social impact and effect of open data at European level. Since I lived in Brussels, I came into contact with many stakeholders that work on research and innovation. Many people knew the issue of open access to research data and were highly interested to give more information on what they knew or wanted to know the results of this thesis. The attention of the Dutch government and the European Union gave this topic a great insight into multi-level governance.

On the one hand, by attending seminars and talking about open access to research data, the topic became more vivid for me and therefore I wanted to know more about open data. On the other hand, by attending seminars and the different conversations on open data resulted in much more (new) information and more insight in the difficulties of the topic. The complexity of open access to research data makes it more interesting to work on, but also highly intensive by getting familiar with the topic and its correlations with other issues as intellectual property rights. In hindsight, I would initiate additional explanatory interviews to get familiar with the research topic. Only reading about open access to research data was possible, but I think additional interviews would have accelerated my knowledge of open access to research data. Moreover, the complexity and the wealth of data may affect this study negatively, because structuring and linking data can therefore be suboptimal. In addition, information from the analysis could unintended be less emphasised in the conclusion due to the wealth of information.

Linking data and answers of experts took a lot of time, more than I had expected in advance. Furthermore, the responses of experts was way less than expected. The input I received was often extensive, but it took much effort to receive input from experts. Additionally, some experts did not give as much as information as preferred, which constrained the interpretation of answers. This should be taken into account by reading this study.

As mentioned before, this study has been structured differently than ordinary theses. I expected to do an 'ordinary' structure, without thinking about the possibility of other thesis structures. Therefore, the switch to a deviating thesis structure took more time than I had expected.

After deciding that the thesis structure should be different, I thought I could immediately start with the research process. However, it took a while to get familiar with the topic and the research method. My focus was often based on a regular thesis structure and forming a theory was therefore more difficult than I thought in advance. Furthermore, the method of a Delphi-analysis was very educational for me. By using a Delphi-analysis and establishing a panel of experts I widened my working experience and strengthened my flexibility. However, in my opinion the Delphi-analysis is to some extent insufficient in the debate on open access to research data. This topic asks for more in-depth information and additional questions, especially if expert write statements which are ambiguous. More contact with some experts or additional interviewees could have been beneficial.

The academic and social relevance of this study is high and will remain important in the future. In particular, the combination of opinions, suggestions and academic literature could be beneficial for stakeholders in open access to research data. This combination is, as far as I know, a new approach of open access to research data and is therefore highly relevant for science. Furthermore, the Dutch ministry of education, culture and science has invited me to present the results of this study and the Dutch ministry of economic affairs asked for the results. Both ministries explicitly mentioned that this study could be beneficial for them in their activities on open access to research data. These signals underpin the importance and impact of this issue for society, public administration and science.

All in all, this research was an interesting challenge, which made my perspective wider and my knowledge more extensive. This thesis topic gave me many opportunities and many new contacts which can be useful in the future. The period of research and working in Brussels at European level made a great addition to my academic experience and future career.

6.3 *Recommendations future research*

This study highlighted several aspects of the debate on open access to research data. It is a complex debate, which demands more research on open access to research data. Not many studies have been conducted on open access to research data, therefore this study could not build on earlier research. Sharing these studies and/or data could help future research. Issues that are highly recommended to study are: the feasibility of a regulatory framework, intellectual property rights and its relation with open access to research data. More insight into legislation, legislative barriers, contradictions and specific database rights and copyrights at national and supranational level could be useful for the debate on open access to research data and possible (governmental) regulation strategies.

Furthermore, future research could take a more wider perspective for the Delphi-analysis in order to gain more knowledge on current opinions on open access to research data. Experts of the

Delphi-analysis in this study mainly come from the educational sector/ research institutions. The perspective of publishers and stakeholders in the private sector could give different perspectives during the several Delphi phases.

Future research can take a different focus of open access to research data by focussing on the role of an informed government. To what extent is the government informed about the open access to research data and what can be improved to maintain its nodality position in society in the current digital era.

Furthermore, first (small) conclusions of the pilot of the European Union on open access to research data can be expected at the end of 2014. Future research can focus on the first experiences of stakeholders that participated in the pilot by giving access to their data and stakeholders that choose for an opt-out. Based on this comparison a more in-depth (national and supranational) strategy can be made.

References

- Adler, M. and Ziglio, E. (1996). *Gazing into the oracle: the Delphi method and its application to social policy and public health*. London, United Kingdom: Jessica Kingsley Publishers.
- Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., ... & Wouters, P. (2004). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal*, 3, 135-152.
- Balch, G. I. (1980). The stick, the carrot, and other strategies: A theoretical analysis of governmental intervention. *Law & Policy*. 2(1), 35-60.
- Bemelmans-Videc, M. L. (1998). Policy instrument Choice and Evaluation. In Bemelmans-Videc, M. L., Rist, R. C., and Vedung, E. O. (1998). *Carrots, sticks, and sermons: Policy instruments and their evaluation* (pp. 21-58). New Brunswick, NJ: Transaction Publishers.
- Bemelmans-Videc, M. L., Rist, R. C., and Vedung, E. O. (1998). *Carrots, sticks, and sermons: Policy instruments and their evaluation* (pp. 21-58). New Brunswick, NJ: Transaction Publishers.
- Björk, B., Welling, P., Laakso, M., Majlender, P., Hedlund T and Guðnason, G. (2010). Open Access to the Scientific Journal Literature: Situation 2009. *PLoS ONE*. 5(6): e11273. DOI:10.1371/journal.pone.0011273.
- Black, J. (2002). Critical reflections on regulation. *Australian Journal of Legal Philosophy*, 27, 1.
- Bloem, J., Doorn, M. van, Duivestein, S., Manen, T. van, Ommeren, E. van and Sachdeva, S. (2013). *No more secrets with Big Data Analytics*. Retrieved February 14, 2014, from http://vint.sogeti.com/wp-content/uploads/2013/11/Sogeti_NoMoreSecrets.pdf.
- Boeije, H. (2005). *Analyseren in kwalitatief onderzoek*. Den Haag, the Netherlands: Boom Lemma Uitgevers.
- Bowen, G. (2006). Grounded theory and sensitizing concepts. *International journal of qualitative methods*. 5(3), 12-23.

Buchholtz, S., Bukowski, M. and Śniegocki, A. (2013). *Big and open data in Europe. A growth engine or a missed opportunity?* Retrieved February 7, 2014 from <http://www.microsoft.com/global/eu/RenderingAssets/pdf/2014%20Jan%2028%20EMEA%20Big%20and%20Open%20Data%20Report%20-%20Final%20Report.pdf>.

Canessa, E. and Zennaro, M. (2008). *Science Dissemination using Open Access*. Retrieved February 28, 2014, from <http://sdu.ictp.it/openaccess/SciDissOpenAccess.pdf>.

Charmaz, K. (2006). *Constructing grounded theory*. Thousand Oaks, CA: Sage.

Correia, A. M. R. and Teixeira, J. C., (2005). Reforming scholarly publishing and knowledge communication: From the advent of the scholarly journal to the challenges of open access. *Online Information Review*. 29(4), 349-364.

COST (2013). *Open Access: Practice, Opportunities and Challenges*. Retrieved June 4, 2014, from: file:///C:/Users/Eigenaar/Downloads/COST_Open_Access_Report.pdf.

Dalkey, N. C., Brown, B. B. and Cochran, S. (1969). *The Delphi method: An experimental study of group opinion*, 3. Santa Monica, CA: Rand Corporation.

Data Archiving and Networked Services (DANS) (n.d.). *Persistent Identifiers*. Retrieved September 19, 2014, from <http://www.dans.knaw.nl/content/persistent-identifiers>.

Data Archiving and Networked Services (DANS) (2013a). *A federated data infrastructure for the Netherlands: The front office - back office model*. Retrieved June 28, 2014, from <http://www.dans.knaw.nl/sites/default/files/file/DANS%20FO-BO-model%20UK%20V1%20DEF%20web.pdf>.

Data Archiving and Networked Services (DANS) (2013b). *Datamanagementplan voor wetenschappelijk onderzoek. Tips voor onderzoekers*. Retrieved April, 9, 2014, from <http://www.dans.knaw.nl/sites/default/files/file/Datamanagementplan%20NL%20def%20web.pdf>

Dillo, I. and Doorn, P. (2011). *The Dutch data landscape in 32 interviews and a survey*. Retrieved March 13, 2014, from http://www.dans.knaw.nl/sites/default/files/file/publicaties/The_Dutch_Datalandscape_DEF.pdf.

DIN (2014). *DIN 31644 Information and documentation - Criteria for trustworthy digital archives*. Retrieved July 30, 2014 from <http://www.nabd.din.de/cmd?level=tpl-art-detailansicht&committeeid=54738855&artid=147058907&languageid=en&bcrumblevel=3>.

DOI (2013). *Introduction*. Retrieved September 19, 2014, from http://www.doi.org/doi_handbook/1_Introduction.html.

Dunleavy, P. and Margetts, H. (2000). The advent of digital government: Public bureaucracies and the state in the internet age. In *Paper delivered at the 2000 Annual Meeting of the American Political Science Association, Washington, DC*.

Escher, T., Margetts, H., Petricek, V., and Cox, I. (2006). Governing from the centre: comparing the nodality of digital governments. *American Political Science Association*. DOI: 10.2139/ssrn.1755762

European Commission (2012a). *Towards better access to scientific information: Boosting the benefits of public investments in research*. Retrieved March 17, 2014, from http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf.

European Commission (2012b). *Scientific data: open access to research results will boost Europe's innovation capacity*. Retrieved March 31, 2014, from http://europa.eu/rapid/press-release_IP-12-790_en.htm.

European Commission (2012c). *Recommendation on access to and preservation of scientific information*. Retrieved June 16, 2014, from http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf.

European Commission (2013a). *Open access*. Retrieved February 6, 2014, from <http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1294&lang=1>.

European Commission (2013b). *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. Retrieved February 17, 2014, from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

European Commission (2013c). *ERC Work Programme 2014*. Retrieved April 11, 2014, from http://erc.europa.eu/sites/default/files/document/file/ERC_Work_Programme_2014.pdf.

European Commission (2013d). *Guideline on Data Management in Horizon 2020*. Retrieved July 11, 2014, from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

European Commission (2013e). *Rules for participation and dissemination in "Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020)"*. Retrieved July 18, 2014, from http://ec.europa.eu/research/participants/data/ref/h2020/legal_basis/rules_participation/h2020-rules-participation_en.pdf.

European Commission (2014). *Horizon 2020 Annotated Model Grant Agreements*. Retrieved July 12, 2014, from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf.

European IPR Helpdesk (2014). *Fact Sheet - Open access to publications and data in Horizon 2020: Frequently Asked Questions (FAQ)*. Retrieved May 27, 2014, from https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Open_Access_in_H2020_0.pdf.

Eysenbach, G. (2006). *Citation advantage of open access articles*. *PLoS Biology* 4(5), e157. DOI: 10.1371/journal.pbio.0040157.

Finn R., Wadhwa K, Taylor M., Sveinsdottir T., Noorman M., Wyatt and Sondervan J. (2014). *Deliverable D3.1: Legal and ethical issues in open access and data dissemination and preservation*. Retrieved July 27, 2014, from <http://recodeproject.eu/wp-content/uploads/2014/05/D3.1-legal-and-ethical-issues-FINAL.pdf>

FORCE11 (2014). *Joint Declaration of Data Citation Principles – FINAL*. Retrieved August 13, 2014, from <https://www.force11.org/datacitation>.

García-Peñalvo, F. J., García de Figuerola, C. and Merlo, J. A. (2010). Open knowledge: challenges and facts. *Online Information Review*, 34(4), 520-539.

GÉANT (2014). *About eduGAIN*. Retrieved August 5, 2014, from http://www.geant.net/service/eduGAIN/about_edugain/Pages/AbouteduGAIN.aspx

Ginev, D. (1998). Rhetoric and double hermeneutics in the human sciences. *Human studies*, 21(3), 259-271.

Guibault, L. (2013). Licensing Research Data under Open Access Conditions, in D. Beldiman, *Information and Knowledge: 21st Century Challenges in Intellectual Property and Knowledge Governance*, Cheltenham, Edward Elgar.

Guibault, L. and Wiebe, A. (2013). *Safe to be open. Study on the protection of research data and recommendations for access and usage*. Retrieved March 6, 2014, from <http://webdoc.sub.gwdg.de/univerlag/2013/legalstudy.pdf>.

Hague, R. and Harrop, M. (2010). *Comparative government and politics*. London: Palgrave Macmillan.

Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance*, 21(1), 40-46.

Hood, C. (2007). Intellectual obsolescence and intellectual makeovers: reflections on the tools of government after two decades. *Governance*, 20(1), 127-144.

Hood, C., and Margetts, H. (2007). *The tools of government in the digital age*. Hampshire, United Kingdom: Palgrave Macmillan.

Howe, D., Constanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, . . . Yon Rhee, S. (2008). Big data: The future of biocuration. *Nature*. 455, 47-50. DOI: 10.1038/455047a.

Hsu, C. C. and Sandford, B. A. (2007). The Delphi technique: making sense of consensus. *Practical Assessment, Research & Evaluation*, 12(10), 1-8.

International Telecommunication Union (2014). *ICT Fact and Figures*. Retrieved September 9, 2014, from <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2014-e.pdf>.

ISO (n.d.). *ISO 16363:2012 Space data and information transfer systems -- Audit and certification of trustworthy digital repositories*. Retrieved July 30, 2014, from http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510.

Jordan, A., Wurzel, R. K., and Zito, A. (2005). The rise of 'new' policy instruments in comparative perspective: has governance eclipsed government?. *Political studies*, 53(3), 477-496.

Lemaire, L. (1998). The stick: regulation as a tool of government. In Bemelmans-Videc, M. L., Rist, R. C., and Vedung, E. O. (1998). *Carrots, sticks, and sermons: Policy instruments and their evaluation* (pp. 21-58). New Brunswick, NJ: Transaction Publishers.

Levi-Faur, D. (2011). *Handbook on the Politics of Regulation*. Cheltenham, United Kingdom: Edward Elgar Publishing.

Linstone, H. A. and Turoff, M. (2002). The Delphi Method. *Techniques and applications*, 53.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Hung Byers, A. (2011). *Big data the next frontier for innovation, competition and productivity*. Retrieved March 17, 2014, from http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx.

Max Planck Gesellschaft (2003). *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. Retrieved March 17, 2014, from http://openaccess.mpg.de/67605/berlin_declaration_engl.pdf.

Mcdonald, S. (n.d.). *Research data explained. MANTRA Research Data Management Training*. Retrieved March 12, 2014, from http://datalib.edina.ac.uk/xerte/play.php?template_id=9.

Mitchell, V. W. (1991). The Delphi technique: an exposition and application. *Technology Analysis & Strategic Management*, 3(4), 333-358.

Molloy, J.C. (2011). The Open Knowledge Foundation: Open Data Means Better Science. *PLoS Biology*. 9(12), p. 1-4. DOI: 10.1371/journal.pbio.1001195.

OECD (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Retrieved March 16, 2014, from <http://www.oecd.org/science/sci-tech/38500813.pdf>.

Okoli, C. and Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*, 42(1), 15-29.

Open Definition (n.d.). *The Open Definition*. Retrieved July, 17, 2014, from <http://opendefinition.org/>.

Parker, C. (2007). Meta-regulation: legal accountability for corporate social responsibility. Retrieved September 15, 2014, from http://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID943955_code353054.pdf?abstractid=942157&mirid=1

Peters, B. G. and Van Nispen, F. K. M. (1998). *Public policy instruments. Evaluating the tools of public administration*. Cheltenham, United Kingdom: Edward Elgar Publishing Limited.

PloS (n.d.1). *About*. Retrieved August 7, 2014, from <http://www.plos.org/about/>.

PLoS (n.d.2). *Altmetrics*. Retrieved July 2, 2014, from <http://article-level-metrics.plos.org/alt-metrics/>.

Priem, J., Taraborelli, D. Groth, P. and Neylon, C. (2010). *Altmetrics: A manifesto*. Retrieved September 19, 2014, from <http://altmetrics.org/manifesto>.

Project Open Data (n.d.). *Open Data Policy — Managing Information as an Asset*. Retrieved September 19, 2014, from <http://project-open-data.github.io/>.

Raadschelders, J. C. (2003). *Government: A public administration perspective*. New York, NY: ME Sharpe.

REF (n.d.) *Research Excellence Framework*. Retrieved August 7, 2014, from <http://www.ref.ac.uk/>.

Research Libraries Group (2002). *Trusted Digital Repositories: Attributes and Responsibilities*. Retrieved June 28, 2014, from <http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf?urlm=161690>.

Rijksoverheid (2013). *Kamerbrief over Open Access van publicaties*. Retrieved January 23, 2014, from <http://www.rijksoverheid.nl/documenten-en-publicaties/kamerstukken/2013/11/15/kamerbrief-over-open-access-van-publicaties.html>.

Rowe, G. and Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International journal of forecasting*, 15(4), 353-375.

Schmidt, B. and Kuchma, I. (2013). *Implementing Open Access Mandates*. Retrieved March 12, 2014, from <http://www.oopen.org/download?type=document&docid=436832>.

Schmitt, C. (2013). Security and Privacy in the Era of Big Data. Retrieved February 16, 2014, from <http://www.renci.org/wp-content/uploads/2013/11/0313WhitePaper-iRODS.pdf>.

Somerville, J. A. (2008). Excerpt from Somerville, J. A. (2007). *Critical factors affecting the meaningful assessment of student learning outcomes: A Delphi study of the opinions of community college personnel*. Unpublished doctoral dissertation, Oregon State University, Corvallis. Retrieved March 27, 2014, from <http://jasomerville.com/wp-content/uploads/2011/08/DelphiProcess080617b.pdf>.

Tene, O. and Polonetsky, J. (2012). Privacy in the age of big data: a time for big decisions. *Stanford Law Review Online*. 64, 63-69.

Tene, O. and Polentetsky, J. (2013). Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property*. 11(5), 239-272.

Van Dijk, J. and Landsheer, H. (2011). Delphi-onderzoek en andere methoden voor het raadplegen van deskundigen, in H. Landsheer, H. 't Hart, M. de Goede, M. and J. van Dijk (2003), in *Praktijk gestuurd Onderzoek: Methoden van Praktijkonderzoek*, Groningen: Stenfert Kroese.

UK Open Research (2014). *UK Open Research Data Forum*. Retrieved June 13, 2014, from https://royalsociety.org/~media/Royal_Society_Content/z_events/2014/2014-01-21-Forum-Note.pdf.

Universiteit Leiden (n.d.). *Betrouwbaarheid*. Retrieved September 16, 2014, from <http://www.leidenuniv.nl/fsw/psychologielexicon/index.php3-c=144.htm>.

Van den Hoonaard, W. C. (1997). *Working with sensitizing concepts: Analytical field research*. Thousand Oaks, CA: Sage.

Van der Doelen, F. C. J. (1998). The “give and take” packaging of policy instruments: optimizing legitimacy and effectiveness. In Bemelmans-Videc, M. L., Rist, R. C., and Vedung, E. O. (1998). *Carrots, sticks, and sermons: Policy instruments and their evaluation* (pp. 21-58). New Brunswick, NJ: Transaction Publishers.

Van Thiel, S. (2007). *Bestuurskundig onderzoek. Een methodologische inleiding*. Bussum, the Netherlands: Coutinho.

Vedung, E. (1998). Policy instruments: typologies and theories. In Bemelmans-Videc, M. L., Rist, R. C., and Vedung, E. O. (1998) *Carrots, sticks, and sermons: Policy instruments and their evaluation* (pp. 21-58). New Brunswick, NJ: Transaction Publishers.

Vedung, E. and Van der Doelen (1998). The sermon: information programs in the public policy process – choice effects and evaluation. In Bemelmans-Videc, M. L., Rist, R. C., and Vedung, E. O. (1998). *Carrots, sticks, and sermons: Policy instruments and their evaluation* (pp. 21-58). New Brunswick, NJ: Transaction Publishers.

Vierkant, P. (2012). Visualizing Open Access. *LIBREAS. Library Ideas*, 21. Found on March 25, 2014, from <http://libreas.eu/ausgabe21/texte/05vierkant.htm>

Vu, K. M. (2011). ICT as a source of economic growth in the information age: Empirical evidence from the 1996–2005 period. *Telecommunications Policy*. 35, 357–372.

Appendix A | List of Delphi-analysis experts

Experts in the Delphi-analysis

- **Bruce, Rachel** | Jisc
 - Deputy chief innovation officer at Joint information systems committee (Jisc).
- **De Nooijer, Ivo** | LURIS Leiden
 - Director of Leiden University Research and Innovation Services (LURIS).
- **Dillo, Ingrid** | DANS
 - Director Policy Development Data Archiving and Networked Services (DANS).
- **Madho, Arvin** | RDW and PBLQ
 - Policy Advisor at Government Road Transport (RDW), researcher Open Data and information management at PBLQ.
- **Meyer, Dagmar** | ERCEA
 - Policy Adviser at European Research Council Executive Agency (ERCEA).
- **Ridder-Numan, Jeanette** | Dutch Ministry of Education, Culture and Science
 - Coordinator Large-Scale Research infrastructures, Coordinator Life Sciences.
- **Rombouts, Jeroen** | 3TU.Datacentrum
 - Head Research Data Services at TU Delft and Managing Director 3TU.Datacentrum.
- **Sondervan, Jeroen** | Amsterdam University Press
 - Project Manager at Amsterdam University Press and RECODE project and Project Manager Open Access at Brill.
- **Van Nieuwerburgh, Inge** | Ghent University
 - Coordinator scholarly communications at Ghent University and member of OpenAIRE project.
- **Wittenburg, Peter** | Research Data Alliance
 - Senior Advisor of the Language Archive Unit at Max Planck Institute for Psycholinguistics and member of Research Data Alliance (RDA) Technical Advisory Board.

Appendix B | List of interviewees

Overview of interviewees

- **Grosfeld, Thomas** July 2, 2014, Brussels, Belgium (by telephone)
 - Secretary policy team General Economic policy, Education, Innovation and Health.

- **Hof, Cees** July 3, 2014, Brussels, Belgium (by telephone)
 - General Coordinator at Netherlands Biodiversity Information Facility (NLBIF) & external advisor data management.

- **Kolman, Michiel** July 17, 2014, Brussels, Belgium (by telephone)
 - Senior Vice President Global Academic Relations ELSEVIER.

- **Ridder-Numan, Jeannette** April 15, 2014, The Hague, the Netherlands
 - Board member Research and Scientific policies, Dutch Ministry of Education, Culture and Science. An exploratory interview has been conducted and she participated in the Delphi-analysis.

- **Spichtinger, Daniel** July 14, 2014, Brussels, Belgium
 - Policy Officer at European Commission Directorate-General Research and Innovation.

- **Van den Biesen, Jan** May 20, 2014, Brussels, Belgium
 - Vice President Public Research and Development Programs, Philips Research and responsible for the open access in research and development programmes of Philips.

- **Van Loenen, Dries** May 27, 2014, The Hague, the Netherlands
 - Special envoy International Research, Development and Innovation Electronic Components and Systems, Dutch Ministry of Economic affairs.

Appendix C | Open research data pilot areas

Source: European Commission, 2013b; European IPR Helpdesk, 2014, p. 9.

Seven areas of Horizon 2020 participate in the Open Research Data Pilot. Other individual projects funded under Horizon 2020 and not covered by the scope of the pilot may participate on a voluntary case-by-case basis. The areas:

- Future and Emerging Technologies;
- Research infrastructures – part e-Infrastructures;
- Leadership in enabling and industrial technologies – Information and Communication Technologies;
- Societal Challenge: 'Secure, Clean and Efficient Energy' – part Smart cities and communities;
- Societal Challenge: 'Climate Action, Environment, Resource Efficiency and Raw materials' – except raw materials;
- Societal Challenge: 'Europe in a changing world – inclusive, innovative and reflective Societies'; and
- Science with and for Society.

Appendix D | BusinessEurope position paper

Source: Mr. Jan van den Biesen, Philips. Retrieved from:

http://www.google.nl/url?url=http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm%3Fdoc_id%3D2826&rct=j&frm=1&q=&esrc=s&sa=U&ei=sxa4U5bmCKTd4QThrYCoBw&ved=0CBQQFjAA&usg=AFQjCNEgxMPdCFLCipUrE1pzuGW_nKxGQg

Key messages

1. Open access to scientific information from publicly funded research can stimulate and promote research and innovation activities only if the protection of intellectual property rights is ensured.
2. Open access for scientific publications from publicly funded research must not interfere with commercial exploitation of research results or the possibility to protect results by intellectual property rights.
3. Open access to research data deriving from publicly funded programmes must only apply on a voluntary basis and provided that the legitimate interests of all partners in the consortia are not adversely affected.

Policy developments on open access

On 17 July 2012, the European Commission published a Communication¹ and a Recommendation² to promote open access for scientific information resulting from publicly funded research. BUSINESSEUROPE supports in principle the concept of open access as a driver of scientific and technological progress provided that some essential conditions are fulfilled.

In general terms, the protection of intellectual property rights must always be ensured if open access to scientific information is to stimulate and promote research and innovation activities. Furthermore, it is essential to make a distinction between open access to scientific publications and open access to research data.

Open access to scientific publications: boundary conditions must apply

In the past BUSINESSEUROPE³ has specified the necessary conditions to be respected when promoting open access for scientific publications from publicly funded research:

- the scope must be limited to published versions or final manuscripts accepted for publication resulting from EU-funded or co-funded research projects;

- there must be no obligation to publish; and
- there must be no interference with commercial exploitation of research results or the possibility to protect results by intellectual property rights such as patents.

These conditions seem to be acknowledged by the Commission since it is essentially proposed to extend the current open access pilot running under above conditions within the Framework Programme 7 to the entire Horizon 2020 programme. BUSINESSEUROPE welcomes this approach.

Open access to research data: legitimate interest of all partners to be protected

The Commission is also proposing to promote open access to research data resulting from publicly funded research. On this purpose the Commission will to set up a pilot scheme on open access to and re-use of research data generated by projects in selected areas of Horizon 2020.

The conditions under which the pilot will operate are not defined and wording in the recent Commission's proposals suggests that further clarification is needed. The Commission Recommendation⁴ states that: *"[...] concerns in particular in relation to privacy, trade secrets, national security, legitimate commercial interests and to intellectual property rights shall be duly taken into account. Any data, know-how and/or information whatever their form or nature which are held by private parties in a joint public/private partnership prior to the research action and have been identified as such shall not fall under such an obligation"*.

BUSINESSEUROPE would like to express its serious concerns regarding this formulation. It implies that data, know-how, knowledge and/or information resulting as outputs from publicly funded projects involving private parties are to be disregarded. This is inconsistent with the principles set out above and unacceptable.

Legitimate commercial interests require exemption from any open access obligation to research data. Failing to do so would jeopardise the interest of private parties in participating and investing in such publicly funded projects and therefore reduce the innovation potential of EU research and innovation programmes.

BUSINESSEUROPE urges the EU institutions to address this problem by clarifying that open access to research data deriving from publicly funded programmes must only apply on a voluntary basis and provided that the legitimate interests of all partners in the consortia are not adversely affected.

* 1 COM(2012) 401 “Towards better access to scientific information: boosting the benefits of public investments in research”

* 2 COM(2012) 4890 « Recommendation on access to and preservation of scientific information”

Open access to scientific information in EU funded programmes

* 3 BUSINESSEUROPE, 23-02-2011 “Open Access in the EU’s research programmes”

<http://www.businessseurope.eu/content/default.asp?PageID=568&DocID=28127>

* 4 See note 2, page 6

Appendix E | Philips position paper

Source: Mr. Jan van den Biesen

Open Access to research data

Consultation, July 2, 2013, Brussels

In promoting open access, the Commission's Communication of July 2012 fortunately makes a clear distinction between scientific publications and research data, because the issues at stake are fundamentally different.

Open access to scientific publications is not an issue

The debate on open access to scientific publications originates from the high subscription fees that are hampering access to scientific journals. A pilot for putting articles for free on the internet is currently running in parts of FP7, which is adhering to the following three principles:

the scope of the open access obligation is limited to published versions or final manuscripts accepted for publication resulting from EU-funded or co-funded research projects,

there is no obligation to publish,

there is no interference with commercial exploitation of research results or the possibility to protect results by intellectual property rights such as patents.

If these same principles will basically apply, the extension of the current approach in the FP7 pilot to the entire Horizon 2020 programme deserves our full support.

Open access to research data is a different matter

The basic argument for open access to research data and other results from publicly funded research is that taxpayers should not have to pay twice, first for the research and then again for the results. However, open access to research data could limit possibilities to effectively protect the competitive edge of our research results by means of intellectual property rights or confidentiality, or could allow others to freely use our research results and imitate innovations based on those results. By limiting the exploitation potential of our research results, fully open access might actually destroy more value than it creates.

Instead, we think it should be decided case-by-case whether open access is appropriate, in particular for the research that we are partially funding ourselves. Let me illustrate that with some brief

examples. In the case of unsuccessful clinical trials, widely sharing the negative results - while protecting the privacy of patients - can help prevent unnecessary duplication of research. However, in the case of Key Enabling Technologies, open access to all research data could destroy the competitive edge of project participants over their competitors worldwide. Also in many close-to-market activities this will probably be an issue. Open access to research data should therefore only apply on a voluntary basis, to be decided jointly by the project participants and subject to any restrictions due to the protection of intellectual property, security rules or legitimate commercial interests.

Fortunately, the 2012 Communication from the European Commission is to a fair degree addressing the concerns of industry. The Commission intends to set up a pilot scheme in selected areas of Horizon 2020. These pilots will not apply to projects whose primary aim would be contradicted by making research data accessible. The pilots will also take into account any restrictions that may be needed in order to protect intellectual property or legitimate commercial interests. In this context we suggest to have a close look at the open access policy recently adopted by the Executive Office of the US President (Office of Science and Technology Policy), as this seems a quite workable approach.

Of course we agree that publicly funded research should in the end benefit Europe's citizens. The question is how to best achieve that. Just giving citizens access to all research data will not really help. The most efficient way is to let business play its key role in the innovation chain by translating research results into new products and services in the market place and delivering innovation that matters to Europe's citizens.

Jan van den Biesen

13.014

Appendix F | Guidelines Data Seal of Approval

Source: suggested by one expert in the third session of the Delphi-analysis;
<http://datasealofapproval.org/en/information/guidelines/>

Guidelines Relating to Data Producers:

1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data and compliance with disciplinary and ethical norms.
2. The data producer provides the data in formats recommended by the data repository.
3. The data producer provides the data together with the metadata requested by the data repository.

Guidelines Related to Repositories:

4. The data repository has an explicit mission in the area of digital archiving and promulgates it.
5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.
6. The data repository applies documented processes and procedures for managing data storage.
7. The data repository has a plan for long-term preservation of its digital assets.
8. Archiving takes place according to explicit work flows across the data life cycle.
9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.
10. The data repository enables the users to discover and use the data and refer to them in a persistent way.
11. The data repository ensures the integrity of the digital objects and the metadata.
12. The data repository ensures the authenticity of the digital objects and the metadata.
13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.

Guidelines Related to Data Consumers:

14. The data consumer complies with access regulations set by the data repository.
15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.
16. The data consumer respects the applicable licences of the data repository regarding the use of the data.

Appendix G | Legal and ethical issues

Source: RECODE project. Finn, Wadhwa, Taylor, Sveinsdottir, Noorman, Wyatt & Sondervan, 2014.

Deliverable D3.1:

Draft report: Legal and ethical issues in open access to research data

Ownership of and authority for giving access to (open) data are two different things. Intellectual property rights protect creative works by individuals that are the result of innovation, skill and specialist effort.² This may include music, design, inventions, processes or scientific discoveries, as well as others. Intellectual property rights are comprised of moral rights and exploitation rights. Moral rights include rights such as attribution, respecting the work or remaining anonymous and they are often non-transferrable.³ Exploitation rights include the ability to reproduce, distribute, perform, broadcast or transform materials without permission.⁴ Intellectual property rights are governed by intellectual property laws, and the US, Japan and all 28 European Member States are among the members of the World Trade Organisation (WTO) and have signed up to the 1994 Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement) which seeks to harmonise intellectual property laws across the WTO member nations.

The governing of intellectual property rights in relation to open access to research data references both moral rights and exploitation rights for the researchers who created, collected or curated the data. In relation to moral rights, rights of attribution and respecting the integrity of the original work are implicated. With respect to rights of exploitation, these are related to open access to research data through copyright, database rights, trade secrets, patents as well as rights to reproduce, distribute and transform materials. (However, individuals may waive their exploitation rights or trade them through licensing, which will be discussed in more detail in the next section.) Furthermore, individuals or organisations other

² Korn, Naomi, and Charles Oppenheim, *Licensing Open Data: A Practical Guide*, June 2011 version 2.0. http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf

³ Rodríguez-Doncel, Víctor, Asunción Gómez-Pérez and Nandana Mihindukulasooriya, "Rights declaration in Linked Data", in Olaf Hartig, Juan Sequeda, Aidan Hogan, Takahide Matsutsuka (eds.), *Proceedings of the Fourth International Workshop on Consuming Linked Data (COLID2013)*, Sydney, Australia, 22 October 2013, CEUR-WS, Vol. 1034, 2013, p.3. http://ceur-ws.org/Vol-1034/RodriguezDoncelEtAl_COLID2013.pdf

⁴ Ibid.

than researchers themselves may claim “neighbouring” or “related” rights if they have curated the data or translated it in some way.⁵

Copyright, or parts of it, can be transferred or assigned to other persons and or parties. For example a researcher who is working for an university has copyright on their research output (publications) but is likely to transfer (all or parts of) their copyright to a publisher, which will publish his article in a scientific journal. There are many other forms of licensing (e.g. exclusive, non-exclusive or compulsory licensing), which arranges copyright transferring for certain aspects in the distribution and dissemination chain.

Database rights are often relevant to data sets, whether in respect of intellectual property rights or not. Article 10 of the TRIPS agreement includes the following provision in relation to compilations of data. “Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such.”⁶ Database rights are independent of copyright. The arrangement, selection, and presentation of the data may be protected by copyright, while the database, as a whole, can be protected by database right. Thus, database protection does not extend to the data or the material itself. Simple collections of data do not count as intellectual property; it is at the point of organisation and selection that intellectual property rights are recognised. In Europe, a specific database right law, the 1996 Database Directive, protects the producer of a database, who has invested the necessary effort to constitute the database.⁷ Database rights under the EU are created automatically, vested in the employers of creators (when the action of creation was part of employment), and do not have to be registered to have effect.

In relation to open data, database rights prevent third parties from publishing, distributing and copying protected research data. Some of the re-use restrictions claimed by private companies in the Earth Sciences case study are based upon database rights, as the JRC does not own the data it is using for their research. In his discussion on releasing public sector data, De Vries points out that public sector bodies collect information and data sets that they do

⁵ Ibid.

⁶ World Trade Organisation, *Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement)*, 15 April 1994.

⁷ Marc de Vries, *Open Data and Liability*, European Public Sector Information Platform Topic Report No. 2012 / 13, December 2012.

not necessarily own; for example data produced by third parties as a result of research or other contracts.⁸ This leaves public sector bodies vulnerable to legal action by the rights holders. Therefore, if a public body does not hold all of the intellectual property rights associated with the data, it may not be entitled to open up the data for re-use and may need to refrain from doing so.⁹ The Revised PSI Directive recognises this barrier and advises public sector bodies not to release information which third parties hold intellectual property rights under the Directive.¹⁰

⁸ Ibid., p. 7.

⁹ Ibid.

¹⁰ European Commission, Directive 2013/37/EU Amending Directive 2003/98/EC on the re-use of public sector information, *Official Journal of the European Union*, L175, 26 June 2013, pp. 1-8.

Appendix H | Interview topic list

The below given research questions have been used in several interviews. However, not all the used questions are added.

English questions

- What are your responsibilities in the open access to research data process?
- To what data needs to be access?
- Only about publicly funded research or in future also pure business research data? Voluntary?
- Could you elaborate more about the Pilot of Horizon 2020?
 - What are the experiences?
 - What are difficulties?
 - When and how will it be evaluated?
 - Can we expect that this will be implemented?
- What does de European Commission mean with publicly funded research?
- To what extent needs de EU coordinate the open access to research data process?
- What role do the national agencies and authorities need to take up?
- Should incentives be given in order to stimulate the open access to research data process?

Dutch questions

- Wat houdt open access tot wetenschappelijke data voor jullie in?
- Hoe gaan jullie om met open access tot wetenschappelijke data? Biobanken?
- In hoeverre is open access tot wetenschappelijke onderzoeksdata gereguleerd of vastgelegd?
- In hoeverre door de overheid gefinancierd? Ook bedrijfsleven?
- Hoe kan data gevalideerd worden?
- Waar zou de data geplaatst/opgeslagen moeten worden?
- Hoe zou het systeem onderhouden kunnen worden en welk systeem is bruikbaar?
- Wie betaalt het onderhoud van het datasysteem/dataopslag?
- Welk licentie systeem is aan te bevelen?
- Hoe zou het licentie systeem toegepast kunnen worden op (gevoeligere) data?
- Hoe zou data management moeten plaatsvinden?
- Wat zou de rol moeten zijn vanuit de overheid op het gebied van open access tot wetenschappelijke data?

- Wat is hiervan de huidige situatie?
- Hoe verloopt de internationale samenwerking?
- Wat zijn de ervaringen met de pilot onder H2020?
 - Waar loopt men tegenaan?
- Wat zijn legitieme redenen op voorhand om data niet vrij te geven?
 - Hoe zou dit ingebed kunnen worden binnen een eventueel overheidskader?
- Wat zien jullie als bedrijf als 'in achtneming van IPR'?

Appendix I | Delphi-analysis questions

The following questions have been used in the Delphi-analysis.

First session

1. What is the problem for which open access to research data is the solution?
 - a. Keep in mind: what needs to be freely accessible?
2. Is it important that scientific data needs to be accessible to the public (each individual)?
 - a. Please take into account to *whom it is important*, and;
 - b. *Who* needs to provide open access?
3. What *difficulties* and *opportunities* does the process of open access to research data confront?
 - a. Be as precise as possible by giving the context and comprehensive arguments for both aspects.
4. What needs to be the framework (and criteria) for open access to research data?
5. Is there a need for intervention in the open access to research data process in order to realise innovation in scientific research?
 - a. If so, how and by whom?
 - b. If not, why not?
 - c. Is there intervention needed in certain parts of the process of open access to research data or certain research disciplines?
6. To what extent are knowledge and good practices available of open access to research data in other (EU) countries?
 - a. Please specify *those* knowledge and good practices, and;
 - b. Describe your *experiences* with those issues.

Second session

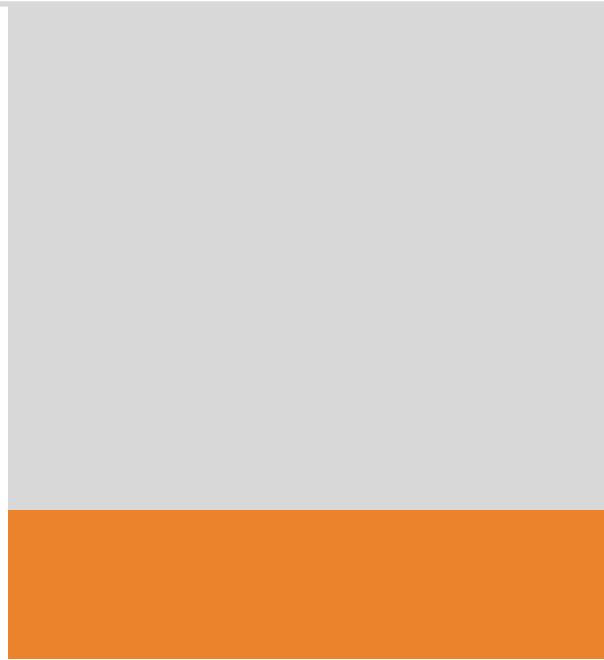
1. There is no consensus on who is or should be the owner of the data and who has the authority to give open access? Please give your opinion about this issue and define the owner of the data. Please take into account the intellectual privacy rights.
2. *Please read page 13-16 of the input appendix carefully.* The input from the first session gave many aspects for a framework, but this remains to general. Please give potential applications and implementation which can be used in the practice. Please read page 20-23 carefully. Specifically comment:
 - a. How can the data management be structured, what needs to be taken into account?
 - b. Where do we need to stall information/data?
 - c. Who needs to check/verify information before it will be accessible?
 - d. Who will maintain and finance the database and its system?
 - e. What are sensitive data or how can we define if data is sensitive?
 - f. Which licensing scheme is necessary?
 - g. How can the security of data be embedded?
 - h. What is a research community?
3. Many participants speak about open access to research data from publicly funded research, but this may constrain the collaboration of R&D in the private sector.
 - a. What is the definition of publicly funded research. Where should you draw the line for publicly funded research which data needs to be accessible? How to distinguish or identify both sources of research?
 - b. How can the private sector and industry be involved in order convince them to give open access to their research data and to share their data?
4. What kind of incentives needs to be given in order to convince researchers to give open access to their research data and to share their data?
5. What will be the consequences for future research by giving open access to research data?

Third session

1. “Ideally research data is raw data, methodology, processes, and processed data”. Do you agree on this? (See page 1 from the output; What is research data?)
2. *Please read page 21-30 of the input appendix carefully:* What will be the criteria to choose the right and appropriate data storage? (e.g. think about national, international, governmental and institutional repositories)
3. *Please read page 31-33 of the input appendix carefully:* Is the proposed governmental role feasible? (What should be its role?)
4. Do you think that the proposed framework could be implemented? (see the input document)
5. Is there (essential) information missing?
6. After reading the input of the second session; do we as society (and its stakeholders in the broad sense) really want to have open access to research data, taking into account the opportunities and difficulties?
7. What if there is no open access to research data?
8. Do you have any final remarks?

coming together is a beginning;
keeping together is progress;
working together is success.

– Henry Ford –



© S.L. Meijer
Brussels - Belgium | 2014
Radboud University Nijmegen | The Netherlands