Building a model for virtual collaboration readiness

Determining what factors affect the attitude of Life Science researchers towards Virtual Research Environments

Éva Kalmár

Master thesis

MSc Science Education and Communication

14. 03. 2016

TU Delft
Delft University of Technology
Science Education and Communication

Building a model for virtual collaboration readiness

Determining the key factors that affect the attitude of Life Science researchers towards Virtual Research Environments

Éva Kalmár

Student number: 4318471

14. 03. 2016

Master Thesis

MSc Science Education and Communication

Contact

Supervisors

Caroline Wehrmann, SEC

Steven Flipse, SEC

Prof. Marc de Vries, SEC

Prof.dr Pieter de Vries, TPM

Company

Biotecture

Nadine Bongaerts

TU Delft
Delft University of Technology

Science Education and Communication

Acknowledgements

One and a half years ago, I was at the Border Sessions Festival in Den Haag as a science communication student, when I was listening to a fascinating discussion between Koert van Mensvoort, the creative director of Next Nature and Nadine Bongaerts, one of the founders of Biotecture. Nadine was interviewing Koert about one of his latest projects, the In Vitro Meat Cookbook, and discussed related scientific technologies, like cloning, tissue culture and large-scale production, but also design, which could help the acceptance of this hypothetical product, the in vitro meat. This discussion was a great example of good science communication. Later on, I contacted Nadine, as she offered me an internship project: to participate in the organization and realization of science outreach training in the EpiPredict consortium. At the consortium's Kick-off meeting the scientific partners had a discussion about where to store their data, how to share them and so on. This raised my attention, and I started to look up collaboration-related platforms, which could help scientists in data management and project management issues. This is how the idea of the research was born. I have to thank all of them for giving me inspiration throughout our discussions and for their help, but especially to Nadine Bongaert, Dr Dirk Stemerding, Dr Pernette Verschure, Prof. Antoine van Kampen, Dr Luca Magnani, Prof. Marianne G. Rots and Dr Alex Michie.

I am also thankful for all those people who agreed to spend their precious time and participated in the discussions and interviews, scientists from the EpiPredict consortia and from the we-NMR. I am grateful that I could talk to Annamaria Carusi, Gwendolyn Kolfschoten, Jan Bot, Matthew Dovey, Thorsten Reimer, Rob Speekenbrink, Rob Hooft, Silvia Olabarriaga, Marcel Reinders, Gera Pronk, Karianne Vermaas, Alexandre Bonvin and Zhiming Zhao, I have learned a lot from our discussions. I am also grateful to all those scientists who filled out the online survey and helped me to get a better overview of how life science researchers work in and think about collaborations.

I cannot thank enough Caroline Wehrmann, my first supervisor, for the endless meetings, discussions and support, and Steven Flipse, my second supervisor for the inspiring discussions. I am thankful to Prof. Dr Pieter de Vries for accepting to be in the examination committee and for his motivating comments.

At last, but not least, I thank my friends and family for their support, patience and trust in me, and especially Péter, my husband for his endless support.

Leiden, 14.03.2016.

Éva Kalmár

Executive summary

Collaborations are essential for modern scientific research. Research groups cannot perform valuable scientific inquiry without working together with other researchers, due to the growth of knowledge, high specialization of scientific domains, quickly changing technology, and the appearance of complex problems. Individual scientists or single research groups do not possess enough knowledge, expertise and time to perform research which is appreciated by the funding agencies and academic publishers (Hara, Solomon et al. 2003). Nowadays research funding agencies, like the European Science Foundation prefer interdisciplinary, international and inter-institutional collaborations (Sonnenwald 2007), because these have been shown to increase the quality of research, contribute to the growth of scientific knowledge (Wray 2002).

Natural sciences and especially life sciences faced a drastic change in data management, analysis and sharing practices due to the widespread use of inquiry methods generating terabytes of data. This change is represented in the appearance of specific tools designed for life scientists to handle scientific workflows or to help computation of data (De Roure, Goble et al. 2007, Goecks, Nekrutenko et al. 2010).

Virtual Research Environments (Collaborative Virtual Environments, Collaboratories, Cyberinfrastructure, e-Infrastructure, Collaborative e-Research Communities, or Virtual Research Communities) are innovative, online, community-oriented, flexible, and secure working environments designed for scientific research groups working together (Candela, Castelli et al. 2013). VREs have the potential to change research practices, make the academic research faster, more efficient and even more transparent (Junge 2007), or even speed up the shift between fundamental research and applied science, and improve social and economic well-being (Dutton and Jeffreys 2010). Therefore, it is not surprising that research policy makers at institutional or European levels would like to develop and employ virtual research environments (Dutton and Jeffreys 2010, EU retrieved at 28-04-2015). But scientists in life sciences think differently about this topic (Allan 2009). While the EU is investing into the development of monumental VRE projects and universities build their environments separated from each other, scientists would like to use their legacy systems. Some research suggests that what they need is access to data storage and computational resources (including grid computing); as well software and services, but they do not need VREs (Allan et al. 2006).

To investigate this discrepancy between science policy makers and life science researchers, I have performed an extensive literature review, did interviews with VRE researchers, developers and users, and performed an online survey filled in by potential and actual VRE users. By generating a theoretical framework from the literature and related theories, I could rank the factors found throughout the research that determine the attitude of life science researchers towards VREs and to build a virtual collaboration readiness model. This model illustrates the requirements that need to be fulfilled to reach to a state, in which all the circumstances are given to help scientists from life sciences to adopt to virtual research environments. Finally, detailed recommendations to distinct stakeholders are listed in the Discussions.

Table of Contents

1 Introduction 13
1.1 Problem statement 14
1.2 Research goal 14
1.3 Research question 15
1.4 Research sub-questions 15
1.5 Perspective and limitation of the research 15
1.6 Structure of the thesis 16
2 Background 18
2.1 Scientific collaborations 18
2.2 Science and e-science 21
2.3 Virtual research communities and environments 24
2.4 Concepts of the theoretical framework 30
3 Methods 32
3.1 Literature search 32
3.2 Interviews with VRE researchers and developers 33
3.3 Interviews with scientists 33
3.4 Survey 34
4 Literature review 36
4.1 Collaboration readiness 36
4.2 Theoretical framework 48
5 Results of the interviews performed with VRE developers 50
5.1 Methodology 50
5.2 Collaboration infrastructure readiness 51
5.3 Collaboration technology readiness 56
5.4 Collaboration readiness 58
5.5 Conclusions 62
6 Results of the interviews performed with life scientists 64
6.1 Designing the interview questions 64
6.2 Collaboration infrastructure readiness 68
6.3 Collaboration technology readiness 70

6.4 Collaboration readiness 74
6.5 Conclusions 75
7 Results of the survey 77
7.1 Designing the survey 77
7.2 Collaboration infrastructure readiness 81
7.3 Collaboration technology readiness 83
7.4 Collaboration readiness 91
7.5 Conclusions 95
8 Conclusions 99
8.1 Collaboration infrastructure readiness 99
8.2 Collaboration technology readiness 101
8.3 Collaboration readiness 106
9 Building a model for virtual collaboration readiness 110
10 Discussion 114
10.1 A critical view of the research 115
10.2 The adoption of VREs is affected by multiple factors 116
10.3 Science Communication aspects of the research 117
10.4 EU-wide VREs 119
10.5 Community aspects of the research 120
10.6 Guidelines and recommendations based on the virtual collaboration model 122
References 124

List of Figures

Figure 1. The structure of the thesis 16

Figure 2. The research design 17

Figure 3. The five schools of e-research, based on Bartling and Friesike 2014 23

Figure 4. Virtual community types (based on Bos et al. 2007) 24

Figure 5. Important functionalities a VRE should provide (based on Allan 2009) 27

Figure 6. Technology Acceptance Modell (based on Davis et al 1989) 30

Figure 7. The building blocks of the theoretical background 48

Figure 8. The Theoretical framework of virtual collaboration tool readiness 49

Figure 9. Coding tree for collaboration infrastructure readiness dimension 67

Figure 10. Coding tree for the collaboration technology readiness dimension 67

Figure 11. Coding tree for the collaboration readiness dimension 68

Figure 12. Scientists are not sure whether VREs could be beneficial 81

Figure 13. Life scientists' attitude towards an EU-wide VRE 82

Figure 14. Communication tools used in a collaboration setting 83

Figure 15. The result of the cross table participating in science outreach versus role in the collaboration 84

Figure 16. Scientists find networking important in setting up collaborations 84

Figure 17. The attitude of life science researchers towards social media 85

Figure 18. The age-distribution of the attitude towards social media 85

Figure 19. The majority of scientists use social media for personal purposes 86

Figure 20. Half of those scientists who said they did not use social media used SNSs designed for academia 86

Figure 21. Collaboration leaders tend to be more active academic SNS users than participants 87

Figure 22. Scientists using social media for personal or academic reasons had more positive attitude towards VREs 87

Figure 23. Only one out of ten scientists uses project management of workflow management tools (presented in number of cases) 88

Figure 24. Collaboration leaders find it more important to use project manager tools 88

Figure 25. Scientists who use project management tools have intrinsic motivations for collaborating 89

Figure 26. The ways scientists save their data are versatile 89

Figure 27. Data sharing practices of life scientists 90

Figure 28. Less than half of the scientists use metadata 90

Figure 29 Scientists who use workflow systems, all use metadata when saving data 90

Figure 30: Researchers who consider their collaboration as a virtual research group find VREs more useful 92

Figure 31. Scientists in international collaboration see more likely their collaborations as a virtual research group 92

Figure 32. Scientists' motivations for collaboration 93

Figure 33. The majority of the collaboration leaders are in integrative collaborations.......... 94

Figure 34. Scientists in integrative collaboration find it more important to save data for later use or validation purposes 94

Figure 36 The virtual collaboration readiness model 113

List of Tables

Table 1. Keywords used for literature search 32

Table 2. ICT tools used in a VRE (based on Allan 2009) 38

Table 3. Type of researchers based on their activity on academic SNSs (based on Nentwich et al 2012) 41

Table 4. Project management success factors in scientific collaborations (based on Beukers 2011)... 42

Table 5. Levels of data management (based on Kaizer and Kuipers 2014)......................... 43

Table 6. Question types used in the semi-structured interviews 64

Table 7. Interview questions... 65

Table 8. The survey questions 80

Table 9. What scientists want to change regarding communication in a collaboration 83

Table 10. Scientists' definition of collaboration.. 91

Glossary and Abbreviations

Data	arising from observation (non-repeatable), an experiment (repeatable) or a computer stimulation (calculated)
Information	relationship between items of data
Knowledge	understanding the causality in relationships
Metadata	data about data, it can be searched, accessed and reused
Provenance	the origin of the data
ICT	Information and Communication Technology
VRE	Virtual Research Environment
Web2.0	World Wide Web sites with user-generated content, usability, and interoperability

1 Introduction

Collaborations are crucial components of modern scientific research. Due to the growth of knowledge, high specialization of scientific domains, quickly changing technology, and the appearance of complex problems make it almost impossible to research groups to perform their scientific research on their own. Individual scientists or single research groups do not possess enough knowledge, expertise and time to perform research which is appreciated by the funding agencies and academic publishers (Hara, Solomon et al. 2003). Nowadays research funding agencies, like the European Science Foundation prefer **interdisciplinary, international and inter-institutional** collaborations (Sonnenwald 2007). The increasing trend of collaboration also appears in the increase of academic publications with more authors. Scientific articles appear in journals specified for given scientific fields, and they publish the organisational affiliations of the authors. Therefore bibliometric studies can easily analyse trends in co-authoring papers. International co-authorship almost doubled between 1981 and 1995 in all science fields all over the world. (Chompalov 2014). Not surprisingly, as collaborations have been shown to increase the quality of research, contribute to the growth of scientific knowledge (Wray 2002).

With the widespread use of techniques generating enormous amounts of data, like microarrays, DNA sequencing and genome-, proteome- or even microbiome-wide assays, in the past decades **life sciences** faced a drastic shift in data management, data analysis and data sharing practices. This shift was also represented in the appearance of specific tools specially designed for life scientists to handle scientific workflows or to help computation of data (De Roure, Goble et al. 2007, Goecks, Nekrutenko et al. 2010).

Although collaboration became a must in science, it requires special skills from scientists to share knowledge, resources and commitment, to build social capital and trust. Therefore scientific collaborations are not *per se* successful (Hara, Solomon et al. 2003). Differences in working and thinking patterns, expectations and motivations could challenge knowledge sharing and working together (Sonnenwald 2007). Several projects were and are focusing on developing tools that assist and promote collaboration between scientists. **Virtual Research Environments (VREs)** are innovative, online, community-oriented, flexible, and secure working environments designed for scientific research groups working together (Candela, Castelli et al. 2013). As Christopher Brown from JISC said:

> *"A key characteristic of a VRE is that it facilitates collaboration amongst researchers and research teams, providing them with more effective means of collaboratively collecting, manipulating and managing data, as well as collaborative knowledge creation."*

Based on the results of different projects performed over the last twenty years, VREs can speed up the research process by directly bridging distinct steps in the research life cycle, make the research more efficient by using advanced data and project management tools and expand the boundaries of knowledge generation and research methods (Junge 2007). Unfortunately, some of these environments were unsuccessful due to different reasons (Bos, Zimmerman et al. 2007). One of the biggest **problems** was that some of these environments were developed without knowing what scientists use, what they need and what would be useful to them. The existing environments are too general, without the option of being customized in such a way it would be required by scientists. The significant amount of users is also a problem in this situation. As for other technological novelties, most of the potential users need time and reassurance that it is worth to use, but such an environment is not useful if only a minority of the collaborating partners use it. Based on the Technology Acceptance Model (TAM), people's intentions of using a computer system depend on their **attitude**, which is a combination of different factors, like subjective norms, perceived usefulness and perceived ease of use (Davis, Bagozzi et al. 1989).

1.1 Problem statement

As previous research has shown, VREs have the potential to have a significant impact on changing research practices, make the academic research faster, more efficient and even more transparent. Therefore, it is not surprising that research policy makers would like to develop and employ virtual research environments. There are also economic reasons behind this governmental will: VREs could speed up the shift between fundamental research and applied science, and improve social and economic well-being (Dutton and Jeffreys 2010).

With the Horizon2020 einfra-9-2015 call the European Union has decided to invest into the development of virtual research environments. In the call description it was stated that "VREs should integrate resources across all layers of the e-infrastructure (networking, computing, data, software, user interfaces), should foster cross-disciplinary data interoperability and should provide functions allowing data citation and promoting data sharing and trust". Eight projects (EVER-EST, READ, MuG, OpenDreamKit, VRE4EIC, West-Life, BlueBRIDGE, VI-SEEM) were funded and launched following this call. The majority of them are Europe-wide projects.

However, there is a difference between what institutional and European research policy makers (Dutton and Jeffreys 2010, EU retrieved at 28-04-2015) think is needed for scientific collaborations and what scientists in life sciences (Allan 2009) think about this topic regarding the development and usage of e-research tools. While the EU is investing into the development of monumental VRE projects and universities build their environments separated from each other, scientists would like to use their "good old" legacy systems. Based on some research, what they potentially need is access to data storage and computational resources (including grid computing), as well software and services, but they do not need VREs (Allan et al. 2006). This view is strongly arguable, as most scientists are not aware of the new ICT possibilities, or even of the meaning and the potential benefits of a virtual research environment. The lack of knowledge can strongly influence their attitude towards these highly advanced tools.

1.2 Research goal

The aim of this project is to find out what obstacles prevent life science researchers from adopting virtual research environments. The results of the literature review, interviews and the online survey will be used to build a virtual collaboration readiness model. This model will serve as a representation of the current situation, and based on the results, can give recommendations to research institutions, science policy makers, VRE developers and scientists how to change these factors and achieve a state of readiness for adoption.

1.3 Research question

What are the obstacles that prevent life science researchers in the adoption of the virtual research environment and how can VRE developers, research institutes, science policy makers and life science researchers overcome these obstacles?

1.4 Research sub-questions

1. What key factors did previous studies find that determine life science researchers' attitude towards VREs?
2. What do VRE researchers and developers think about the relevance of the factors found in the literature?
3. What is the attitude of life science researchers towards VREs and what are the obstacles they experience that prevent them from adopting VREs?
4. How can VRE developers, research institutes, science policy makers and life science researchers overcome the obstacles?

1.5 Perspective and limitation of the research

In the process of designing and interpreting the research, I use the perspective of communication, more specifically science communication. I am interested in the communications practices utilized in a collaboration setup, in particular in computer-mediated communication means and social media. As virtual research environments include some aspects of social networking sites, the acceptance of these tools is particularly involved in the study.

Keeping communication as a central focus, I use the perspective of the pragmatic school of e-science throughout the research process, while in most of the scholarly publications the general concept of the infrastructure school is used (Bartling and Friesike 2014). The infrastructure school views information and computer technology (ICT) tools that can enhance research activities (therefore Virtual Research Environments as well) as a technical challenge. Scientists, who are taking this perspective, claim that effective collaboration depends on the available tools and applications. They are focusing on technical aspects and due to this point of view; these projects are problem-centred. In my opinion, the problem of life scientists not adopting VREs lies not only in the technology. Based on the Technology Acceptance model, other factors, like subjective norms and perceived usefulness can determine the attitude of the users. In contrast to infrastructure standpoint, the pragmatic school views ICT as a toolkit by which the knowledge creation, knowledge and data sharing and collaborative work can be made more efficient. This perspective is focusing more on the collaboration as a whole, and not on the tools themselves, and more open to discussion and custom-design.

In most of the literature the term VRE is used to define collaboration-helping small-scale infrastructure, but the EU started to finance projects which are much huger and serve entirely different purposes. The big European projects are aiming to recruit all the researchers who are funded by the EU as members and to encourage them to upload their published results for data reuse. Joining this type of VRE can also lead to future collaborations of course, but the major goal is different from the objective of the small VREs.

Therefore, I make a distinction between small-scale and large-scale VREs. In this project, I focus my investigation on the small-scale environments, which are indeed international and only take an outlook on the EU-wide projects, which is one of the limitations of this work. The second major limitation is that I have restricted my research to life science researchers, and the results of this study are probably not generalizable for other disciplines.

1.6 Structure of the thesis

Introduction	•Chapter 1 Problem statement, research question •Chapter 2 Summary of background knowledge •Chapter 3 Methods
Results	•Chapter 4 Literature review, theoretical framework •Chapter 5 Key factors from the VRE developers' perspective •Chapter 6 Key factors from the users' perspective - interview results •Chapter 7 Key factors from the users' perspectives - survey results
Conclusion Discussion	•Chapter 8 Conclusions •Chapter 9 The virtual collaborational readiness modell •Chapter 10 Discussions
References	•Literature used in the document
Appendices (separate document)	•Interview transcripts •Survey results

Figure 1. The structure of the thesis

The detailed structure of the thesis is summarized in **Figure 1**. After this introduction, Chapter 2 concludes relevant background knowledge on the subject of collaboration, e-science and scientific research, trends in e-science to enable to put the findings of the literature review into context. Using the perspective of communication and the perspective of the pragmatic school of e-science I introduce the research infrastructures specifically designed to help collaborations. I also introduce the concepts used later on in the theoretical framework. The methods used to answer the sub-questions, including literature search, interview protocols and methods related to the survey are listed in Chapter 3.

The results of the research are divided into four distinct chapters for the better overview. The research was designed in such a way, that the answers received to the different sub-questions lead to the development of the following phase of the study. **Figure 2.** summarizes **the research design** in details. To answer the sub-questions and the research question, I used quantitative and qualitative research methods. To answer sub-question 1, I performed a detailed literature search through diverse scientific domains (including Information and Communication Technology, Computer-mediated Communication, Research Policy, Technology Management and Sociology). I present these in Chapter 4.1. Based on the problem statement and the findings of the literature search, I have built a Theoretical framework from different theories and models to give a structural basis for the research and to help me finding a relevant answer to the research question. This theoretical framework will be discussed in detail in Chapter 4.2. To answer sub-question 2, I have interviewed experts developing or working with e-infrastructure, especially VREs. The results of these interviews and the conclusion of these can be found in Chapter 5. The key factors found in the literature and the information gained from the interview gave the basis for designing questions to perform semi-structured interviews with life science researchers to answer sub-question 3. The information gained from these interviews and their conclusions are in Chapter 6. The qualitative analysis of the interview answers helped me to design the online questionnaire, which was designed to answer also sub-question 3. The results of this questionnaire, which were analysed quantitatively, can be read in Chapter 7.

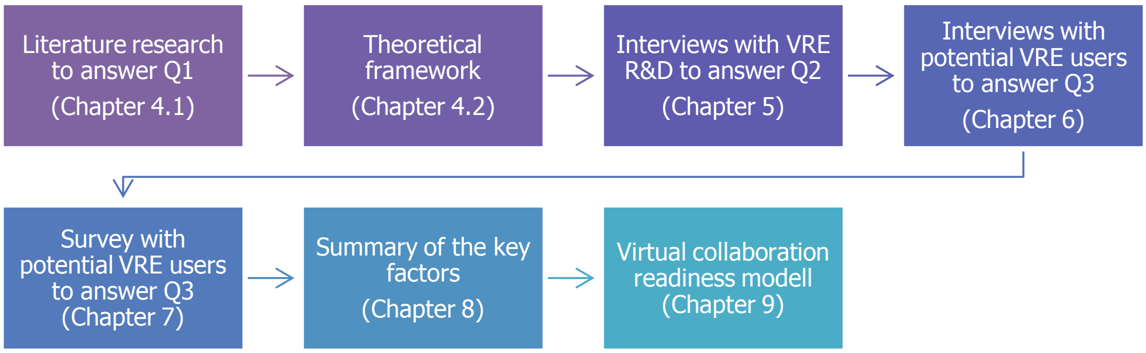

Figure 2. The research design

Chapter 8 contains the Conclusions of all the results, providing answers to the research sub-questions and finally to the research question. In Chapter 9 I introduce the collaboration readiness model, which is based on the results of the entire research. In Chapter 10 I discuss the findings of this research in relation to the previous findings. At the end of the document, there is the list of References, containing all the literature used for making this thesis. The Appendices, containing all the interview transcripts and the original survey results are presented in a separate document.

2 Background

Scientific collaborations are different from collaborations founded in the business world. Chapter 2.1 highlights the specific issues of scientific collaborations, and all those aspects, which will be mentioned later on in the thesis. Chapter 2.2 is focusing on what e-research is and how it can be related to classical scientific research. It also gives an overview of the current schools of e-research, as I came across these topics in the interviews and I found it necessary to make these distinctions. I also highlight the perspective I took when summarizing the literature and then designing the research. Virtual Research Environments belong to the Information and Communication Technology tools of e-research. Chapter 2.3 gives a summary of the tools that are used in collaborations in general and then introduces Virtual Research Environments. Finally, Chapter 2.4 introduces the concepts which are applied in the theoretical framework presented later on in Chapter 4.

2.1 Scientific collaborations

By definition, scientific collaboration is an

> *"interaction taking place within a social context among two or more scientists that facilitates sharing of meaning and completion of tasks with respect to a mutually shared, superordinate goal" (Sonnenwald 2007).*

International collaboration refers to a collaboration between partners from different countries, whereas interdisciplinary collaborations integrate researchers from various scientific domains (Qin, Lancaster et al. 1997). Scientific collaborations are different from other types of collaborations, as tasks often have a high degree of uncertainty due to the nature of scientific research, so the production of one particular outcome or result cannot be promised. Moreover, scientific collaboration is shaped by the social context of scientific research, including scientific paradigms, research policies, measures of success, peer reviewing, and so on (Sonnenwald 2007).

Although collaborations are essential for today's research, scientific inquiring was not always international or multidisciplinary before. It is a result of an increasing trend. International co-authorship almost doubled between 1981 and 1995 in all science fields all over the world. (Chompalov 2014). Collaboration on one hand has affected research positively in many ways, that it reinforced scientists and funding agencies that collaborations are desirable in science. Thagard mentions the example of the bacterial theory of ulcer development, which was only possible to formulate when different medical professionals joined their perspectives in collaborative work (Thagard 1997). Collaborations, in general, lead to reliable, more successful, faster and more efficient knowledge creation (Thagard 1997). On the other hand, one cannot forget the potential negative aspects. Collaborations can hide unethical conduct, can be used to perform intellectual espionage (Sonnenwald 2007), or can lead to diffused responsibilities, especially in ethically sensitive research projects (Wray 2002).

Although some scientists still have a substantial fear that their scientific competitors take advantage on them, sharing scientific data and knowledge between collaborating partners became a norm in most disciplines of natural sciences. Nowadays research funding agencies, like the European Science Foundation prefer to give support to research grants of interdisciplinary, international and inter-institutional collaborations (Sonnenwald 2007). The general reasoning behind that phenomenon is that individual scientists or single research groups do not possess enough knowledge, expertise and time to perform research which is appreciated by the funding agencies and academic publishers (Hara, Solomon et al. 2003). Even fellow scientists appreciate more the results published by multiple authors (Wray 2002).

2.1.1 Stages of scientific collaborations

The life cycle of scientific collaboration is organically attached to the life cycle of scientific research. Collaborations are founded by researchers, who formulate a common goal (in most cases within a joint research project) and then distribute tasks among themselves. After the actual initiation of the collaboration, the whole project has to be sustained, when scientists collect, share and analyse the data and finally at the conclusion stage they publish the results and reflect back to the collaboration itself.

➢ Founding

Why scientists start working together can vary from scientist to scientist and from collaboration to collaboration. The most common reasons are the followings: access to expertise, access to equipment or resources, access to funds, to gain prestige or visibility, to learn tacit knowledge about a technique, to pool knowledge for tackling large and complex problems, to enhance productivity, to teach a student, the increasing specification of science, to encourage cross-fertilization across disciplines and finally for the fun (Bozeman and Corley 2004). Based on Sonnewald (2007), the key success factors for a competent and working collaboration can be grouped into five categories: scientific, political, socioeconomic factors, resource accessibility and finally social network and personal factors (Sonnenwald 2007). From these categories, two cover the focus of this thesis: **scientific factors**, which arise from the trends of the developments in science, like increasing specialization, growing complexity of scientific instruments, the rise of complex scientific topics and **factors linked to social networks and personal issues**. Scientists often neglect this later one, but in fact, they play a crucial role in forming collaborations. Newman (2001) has shown that the small network phenomenon is present in the world of scientific collaborations as well: two scientists are more likely to collaborate with each other if they have been published a paper together before (Newman 2001). Moreover, scientists use their social networks not only for choosing collaborators, but also to gain ideas for new research projects (Sonnenwald 2007).

➢ Formulation

The formulation is the stage of initiating and planning scientific research projects. In the early stages of collaborations, it is crucial to articulate common goals, research visions and to define the tasks, especially the ones regarding leadership. Collaborations are much more successful when the leadership has project management experience and is respected by the team members (Sonnenwald 2007).

➢ Sustainment

As collaboration has been formulated and the actual research has been started, the collaboration needs to be maintained over a period. At this stage, the collaborating partners are focusing on the collection of data, but the relationship between the collaborating peers has to be maintained. The emerging challenges make the sustainability of collaborations sometimes harder than scientists think. Competition between scientists and secrecy are the two major factors that could make the information, knowledge and data sharing difficult. **Communication** is a fundamental part of this stage, too. Without proper communications, tasks could not be synchronised; scientists would not learn from each other, results would not be integrated, and it would be difficult to keep trust. The use of ICT tools can facilitate these communication tasks (Sonnenwald, Solomon et al. 2002).

➢ Conclusion stage

At the final phase, the results emerging from the collaboration are summarized and thesised to the funding agencies. For the evaluation of the collaboration, it is important to define success and remark what the strong and weak points emerged during the whole life cycle of the collaboration (Sonnenwald 2007).

2.2 Science and e-science

Although the stages of collaborations might have been the same 50 years before, the way scientists collaborated was not the same as today's researchers work together. Under research infrastructure then one meant research institutes, computer centers, paper-based laboratory notebooks, file holders and cabinets as data saving possibilities. This view is radically changed. Science infrastructure is taken over by cyberinfrastructure with websites, grids, clouds, automated experiments, digital data generation and analysis, data repositories, blogs and so on (Dutton and Jeffreys 2010). Research is becoming more and more digitalized, as well as scholarly communication. Not only scientific journals are becoming available on the internet, but also, data has to be put into repositories after publishing the results. These data have to be managed, maintained, curated and linked to the publication (Allan 2009).

One important aspect of the digitalization is arising from the fact, that in some scientific disciplines, the amount of data generated is getting huger and huger. In data centred science, data management is a crucial issue, and the majority of the research life cycle is managed by ICT tools.

2.2.1 Data-centred Biology

Due to the appearance of new, large-scale methods and experiments, **life sciences are becoming more and more data-centred**. Based on the existence of this huge amount of data and the costs of the experiments generating these data, it was suggested by several scholars that data should be available for further hypothesis-driven research. However, the reuse of already existing data requires not only extensive and proper data management practices but new attitude from life science researchers as well. Researchers plan and perform experiments without data reuse in mind, and data management and sharing practices are usually not at that level, which could serve the reuse of experimental outcomes. Cyberinfrastructures that enhance data sharing and data stewardship are also favoured by those parties which advocate data-centred research (https://www.surf.nl/en/about-surf, retrieved at 28-04-2015).

In scientific collaborations sharing data, information and knowledge bring up the topic of metadata, data about data. Metadata is usually accessed as a catalogue of information; it contains not only a description, how the data were generated, but also provenance (history of ownership) and location referring to the data set (Allan 2009). These kinds of information are essential for those who would like to use the data uploaded into special data storage places, data repositories.

2.2.2 Data repositories

Data repositories are computer storage facilities where data are kept. The aim of uploading research data to public repositories is to archive them for a longer period of time so that other researchers or potentially anyone from the general public could check the results of a scientific publication. In 2012, the European Commission announced that the Horizon 2020 project will promote and require open access to research data. This means that research projects funded by the Horizon 2020 have to upload the data they produced after publishing (EU retrieved 07.03.2016.). Research repositories are basically defined by the discipline which they serve. These repositories are generally not standardized; they store a broad variety of file formats for access and reuse. There are some initiatives for standardization, but still, data repositories are rarely standardized, which makes it hard to compare or reuse data from different repositories (Pampel, Vierkant et al. 2013).

Digitalization is not only affecting data acquisition and storage but analysis as well. The processes, which happen to data, have to be repeated. Scientific workflow systems were designed to keep track of complicated scientific processes.

2.2.3 Workflow management systems

Workflow management systems are infrastructure or "scientific problem-solving environments" that help in designing, describing and recording complex experiments. A scientific workflow is a defined set of computational or data manipulation steps. These scripts are mostly used in computer-intensive research to inspect and visualize data from different sources. For example in bioinformatics, these workflows cover steps of querying databases, downloading and transforming data and simulations based on the required results (Ludäscher, Altintas et al. 2006). Scientific communities doing similar research can benefit from sharing these workflows with each other, saving time and effort and provide the possibility to compare results with standardized working processes.

Data-intensive scientific research, which generates, stores and analyses huge amounts of data by using computer-mediated tools, become so widespread in the natural sciences, that it received a separate term, e-science.

2.2.4 E-science and e-research

E-science or cyberinfrastructure can be defined as computationally intensive science. It unites different computer technologies which play a crucial role in the generation, storage and analysis of experimental data. Nowadays e-science projects focus more on the long-term archival and access control processes as well (Allan 2009). Cyberinfrastructure has been a US-based project in the 1990`s, and its the basic idea was that knowledge production can be enhanced by computer technologies that help to pool human expertise, data and resources, and provide solutions for the visualisation and computational challenges (Jankowski 2010).

E-research is by definition different from e-science. While e-science is focusing on the computationally intensive science or research generating big data, **e-research** incorporates all the existing and new research that can be enhanced by information and computational technologies. Bartling and Friesike have categorized the recent movements in e-research into five schools, represented **in Figure 3** (Bartling and Friesike 2014). Although not all of them are essential for this project, I have come across most of the issues appearing in the different schools during the interviews made with experts and scientists; therefore, I think it is important to introduce these topics briefly.

The public school (1.) is covering some aspects of the science communication agenda, as it is focusing on the accessibility of the research process and the scientific results as well to a wider audience. Thanks to the new communication technologies, like the web2.0, scientists have much more opportunities to disseminate their findings in a comprehensive manner. The *democratic school* (2.) has more or less the same goals, as it is concerned with the accessibility of scientific knowledge, but the main focus is more on the open accessibility of scientific publications (Open Publication movement) and data (Open Data) produced from publicly-founded projects. *The measurement school* (3.) incorporates those scholars who think that there is a need for alternative standards to measure scientific impact instead of the current impact factor system. The representatives of this school argue that due to the use of web2.0 tools, citation index is not anymore enough to measure the scientific impact, and they suggest the use of altmetrics, which also measures the process of research, collaboration and the communication of the results to the wider audience. *The pragmatic school* (4.) views ICT tools as a toolkit to make distinct research processes more efficient by modularizing knowledge creation, involving external knowledge and helping collaborations with online tools. *The infrastructure school* (5.) sees ICT tools (applications and software tools) as a technical background which enables distant research practices. This school considers Open Science as a "technological challenge", and provides grid computing and virtual environments as solutions for all kinds of research problems occurring in handling big data or in international collaborations.

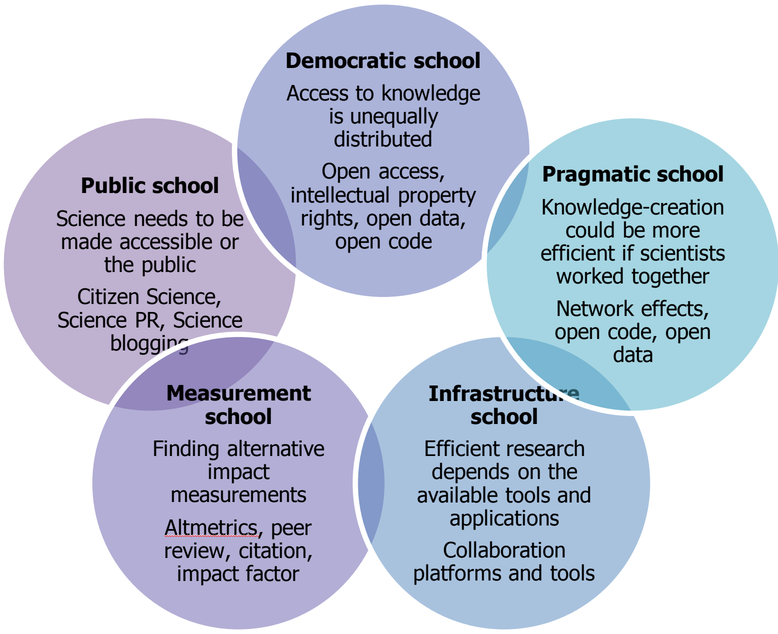

Figure 3. The five schools of e-research, based on Bartling and Friesike 2014

As scientists, who take the perspective of infrastructure school, claim that effective collaboration depends on the available tools and applications, they are focusing on technical aspects and due to this point of view, these projects are problem-centred. In my opinion, the problem of life scientists not adopting VREs lies not only in the technology. Based on the Technology Acceptance model, other factors, like subjective norms and perceived usefulness can determine the attitude of the users. That is why I have chosen to summarize the experiences of the infrastructure school about tools that could help collaborations, but rather take the perspective of the pragmatic school during the research design, in respect of not just focusing on the technical details, but put all of these into the context of collaboration as a whole. Basically, that means that besides the technical perspective I investigate factors which are on the personal or collaboration level.

International collaborations face special challenges; it is hard to sustain these over large distances. Various projects were focusing on the development of ICT tools which can overcome the challenges, including those which were designing VREs, platforms for virtual research communities.

2.3 Virtual research communities and environments

In 1989 Wulf and his group from the Rockefeller University has introduced a new concept, called as *collaboratory,* the name of which arise from merging the words collaboration and laboratory. Their idea was that a collaboratory would be

> *"a center without walls in which researchers can work together regardless of physical location" (Wulf 1993).*

They argued that information and communication technologies were already at that advanced level, that all the activities needed for creating and sustaining a collaboration (from networking to data management) were possible to take place on the internet (Olson, Zimmerman et al. 2008). These ICT-helped collaboration tools could be called as virtual collaborations or virtual research groups (Poole 2009).

2.3.1 Virtual Research Communities

Virtual communities are groups of people communicating through computer-mediated communication channels, as well as via other communication means, like face-to-face interactions or the phone, who all feel that they belong to a certain community. As Rheingold defines them: *"webs of personal relationships in cyberspace"* (Rheingold 1993). If we put this term into research context, then we can define a virtual research community as a group of researchers working together and facilitated by a set of online tools, systems and processes interoperating to support collaborative research **within a virtual research environment** (Allan 2009). The formation of virtual research communities has a beneficial effect on the success of virtual collaborations in all stages (Hossain and Wigand 2004).

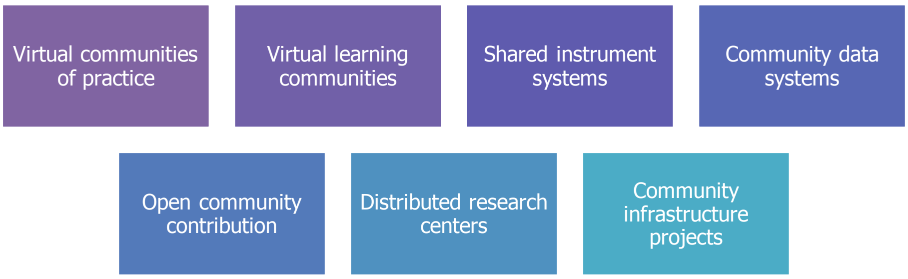

Figure 4. Virtual community types (based on Bos et al. 2007)

Bos, together with his colleagues, have classified **scientific collaborations** based on their ICT-use into seven categories, which are visualized in **Figure 4**. These categories represent different virtual research communities. **Virtual communities of practice** are networks of people who have a common research area and communicate about professional problems, techniques or resources without having a common research project, just sharing information. These are probably one of the best-known community types, besides **virtual learning communities,** which collect people to learn on an online platform. **Shared instrument systems** are based on remote access to some limited, expensive scientific instrument. These are often supplemented with a variety of ICT tools, like video-conferencing, e-lab notebook or chat. **Community data systems** are generated and maintained by specific communities. They serve as information resources, and the information they provide are semi-public and widely used by a given scientific domain. They are usually highly advanced in standardization. The users of **open community contribution systems** have common research problems, and the systems are focused more on work than data. They combine individuals from all over different geographical locations, often from the general public. **Distributed research centers** function like remote research centers: they aggregate scientific talents and resources. These communities have common research projects and shared topics of interest, and most of the communications are personal and informal. **Community infrastructure projects** collect scientists from different disciplines, private sector contractors, ICT personnel who have a common goal: to develop a resource to facilitate science, like software, protocols, educational materials or another kind of infrastructure (Bos, Zimmerman et al. 2007).

Not every internet-based computer-mediated communication forms a virtual community. Virtual groups are only called as virtual communities when they fulfil some criteria. The first one is a **minimum level of interactivity**, which is achieved when all the previous messages in one topic are taken into account when people respond to one particular message. Due to the communication channel, it provides the possibility of spontaneity, interruption, mutuality and turn-taking (Rafaeli and Sudweeks 1997). The second criterion is the **variety of communicators**: there should be more than two people who share messages in the common cyberspace, so database queries and other data-mining activities are not counted as virtual community actions. Virtual communities need a **common public space** where the members of the community can interact with each other, and **a minimum level of the sustained membership** who have a need for communication. The final criterion is the **sense of community,** which unites the feeling of membership, the sense of influence, the need for integration and fulfilment, and the shared emotional connection (Rheingold 1993).

In the case of virtual research communities, the common public space could be provided by **VREs**, which should contain communication tools that enable the minimum level of interactivity. If most members of the user community would use the infrastructure by sharing data, data-related information, and use communication tools to contact other members of the community, the variety of communicators would also be fulfilled. In the case of a small-scale, consortium-size VRE, the community feeling would be more realistic to build, while this could be probably problematic to achieve within a huge EU-wide VRE.

2.3.2 Virtual Research Environments

Virtual Research Environments are by definition

> *"innovative, web-based, community-oriented, comprehensive, flexible, and secure working environments conceived to serve the needs of modern science" (Candela, Castelli et al. 2013).*

The different projects aiming to build an infrastructure to help remote international collaborations run under various names, which became synonyms for VREs: **Collaborative Virtual Environments, Collaboratories, Cyberinfrastructure, e-Infrastructure, Collaborative e-Research Communities, or Virtual Research Communities.** In general, these synonyms mean almost the same, but there might be some differences between these terms. As mentioned in the previous subchapter, virtual research communities or collaborative e-research communities are not infrastructures, but communities, which could use VREs. **Cyberinfrastructure** is a bit misleading term as well, as it is used as the synonym for e-Science, which is a broader term (Allan 2009). **Science Gateways**, as defined at the homepage of EGI (Sciencegateways), are e-infrastructures similar to VREs, but originally these were something different. At the beginning when bioinformaticians started to put algorithms and data analysis tools online, Science Gateways were the portals where these services or software could be reached by internet users. Nowadays their functionality is more or less covers the functionalities of VREs. **Research Infrastructures** are systems that basically use e-Infrastructure to automatize data acquisition, data curation, data access, data processing, and have a community support subsystem as well, which can be a portal or a VRE. This latter one provides the possibility for distinct communities to interact with the data and with each other. From this aspect, Research Infrastructure are not VREs; VREs are only part of them (Chen, Martine et al. 2013).

There are different groups involved in the development, funding and the use of VREs: 1.) those who develop the infrastructure, 2.) governors and managers 3.) the potential user community, 4.) the institutions where the researchers belong to, 5.) the organizations that fund the research projects and finally 6.) the other stakeholders that participate in the research process (Allan 2009). These stakeholders should be taken into account when we talk about different aspects of VREs.

In practice, a small-scale VRE usually has a publicly available home page and a separate site where the authorized members can log in and after authorization, they can use the different ICT tools embedded into the environment. The ICT tools can be analytical tools, data management and community-building tools (Poole 2009). Regarding Allan,

> *a" VREs should not only provide an environment for housing, indexing, and retrieving large data sets but also leverage web2.0 technologies and social networking solutions to give researchers a comprehensive environment for collaboration and resource discovery" (Allan 2009).*

A well-designed VREs could cover the entire research lifecycle, from research administration and project management, through data discovery, collection and analysis, to communication, scholarly publishing and even to career development (Fraser 2005). Data and information uploaded in a Virtual Research Environment are in most cases not yet published or confidential, so a crucial feature of a VRE is confidentiality. With well-designed access control, the members reach only the information they are supposed to access, at the document, library and site or subsite level (SURF 2011).

2.3.2.1 Functionalities

The functions of a standard Virtual Research Environment can be grouped into distinct categories. First of all, to provide the secure and safe environment, VREs should provide federated authentication (with preferably single-sign-on), personalized access control, user management and user statistics. As visualized in **Figure 5**, all the other functions can be built on this base. Document organization and sharing, archives are kind of general tools too, which provide information and data management and sharing. Search facilities help the navigation between information and data. Communication tools, which provide possibilities to communicate synchronously and asynchronously and to share information, are also conventional components of VREs. Project management tools provide help in distribution and completion of collaboration- or research-related tasks.

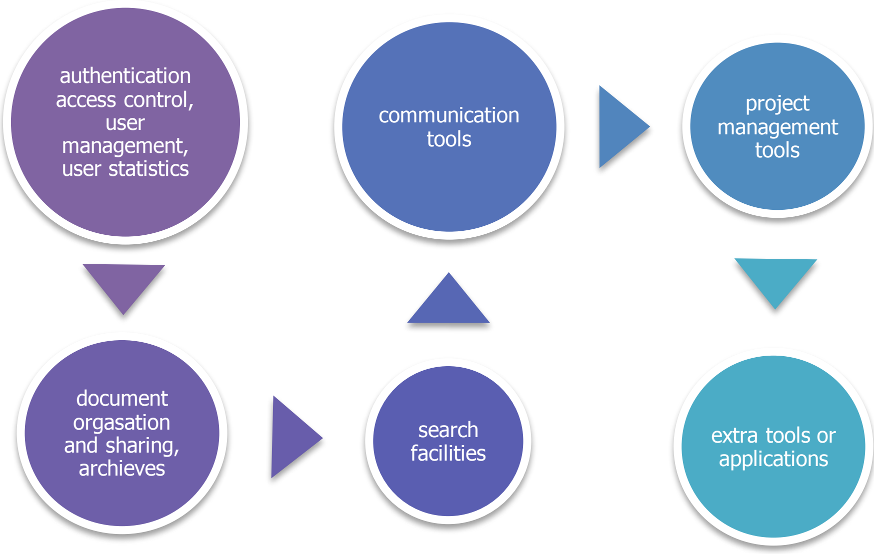

Figure 5. Important functionalities a VRE should provide (based on Allan 2009)

Research registration, alerts, manuscript submission and review systems and reference management are potentially useful extra applications for scientists, as well as the functionality of migrating data to external repositories after publication. Other extra functionalities such as shared research applications, electronic lab notebooks, shared instrumentation can increase the utility of an environment. Finally VREs, especially the bigger ones, can have specialized subsites, for example for different communities, public sites for science outreach or conference sites (Allan 2009).

2.3.2.2 History of VREs

The English Joint Information Systems Committee (**JISC**) has performed extensive research studying virtual research environments between 2004 and 2011. The project had three phases. At the first stage, fourteen projects were funded across the UK, and the aim of this phase was to investigate the definition and explore the possible technical implementations of VREs in research. The basic idea was that virtual research environments could be founded as extensions of virtual learning environments. At the end of this phase, by 2007, it was clear that scientific research needs more specialized implications to use only one framework. These projects found out that VRE developers need to ask scientists about their actual needs and build VREs to satisfy the requests of the customers. The idea of participative design was the basis for the second phase. This two-year –long phase funded four pilot VRE projects in which this participative design would lead the development of VREs fitting the actual collaborations' needs. The third phase started at 2009, and the major goals of this project were to investigate the possibilities of putting specialized ICT tools into a general environment and to develop interoperable solutions (van der Vaart 2010). The final thesiss summarizing the results of the second phase of the VRE project concluded that VREs could speed up the research process by directly bridging distinct steps in the research life cycle, like data collection, processing and analysis, thesis and publication writing. The research processes can be more efficient as well by using more efficient ways of data and project management. Moreover, VREs could provide more transparent and reliable research processes, and could expand the boundaries of the knowledge generation and research methods (Junge 2007).

Collaboratory.nl was a joint project of Novay, Corus, DSM, Phillips, FEI and University of Amsterdam. The research (from 2003 to 2006) was aiming to build an infrastructure for remote operation of lab instruments and remote groupware for remote collaboration between industry researchers and clients. The infrastructure allowed remote experimenting, consulting with researchers, data storage and analysis. The project resulted in a working prototype, which was further developed into commercial products.

The **PARTNER project**, led by the Utrecht University between 2006 and 2009 were building SharePoint-based virtual knowledge centres in research groups.

SURF, the Dutch organisation for ICT for education and research organisations, had a **Collaboration Infrastructure (COIN) project**, which aimed to share collaboration functionality across educational institutions and to provide a generic infrastructure that enables a seamless integration between service providers and service consumers. The project owner, SURFfoundation performed research to see whether the available VREs cover what scientists needed and what factors play a role in the choice and implementations of these tools (van der Vaart 2010). They launched their collaboration platform, SURFgroups in 2006. SURF has concluded that here is a need for an infrastructure, which could offer supporting services (like federated access, group management, messaging tools that work between applications, and so on.) for helping collaborations in a multi-domain environment. They suggested that these tools should be interoperable and placed in a portal-like environment (van der Vaart 2010). Later on, between 2008 and 2011, they have run another project, the SURFshare programme. It aimed to create a common infrastructure which facilitates access to research information and enables sharing scientific information. The project was also financing the Collaboratory project is 2007 and 2008.

2.3.2.3 Examples of life science VREs

The list of VREs in this chapter is less than complete; these were randomly picked to show some versatility of VREs used in the life sciences.

Zfin is a VRE for the zebrafish research community. It is a resource of curated genetic and genomic information related to zebrafish: genes, mutants, genotypes, expression patterns, sequences, thesiser lines and so on. Data are retrieved from publications, but researchers can also directly upload data too. There is a community Wiki for scientists to share information with each other (Ruzicka, Bradford et al. 2015).

Go-Geo is a scientific gateway specialized to geospatial data and related resources. They support standard metadata creation for geographical data (Mathys 2004).

We-NMR is a global e-Infrastructure for NMR and structural biology. The platform integrates and streamlines the computational approaches necessary for NMR data analysis and structural modelling. Access to the e-Infrastructure is provided through a portal integrating commonly used NMR software and grid technology (Bonvin, Rosato et al. 2010).

NCIP Hub is a VRE for the National Cancer Information Program, based on HUBZero. Users from all over the world can share resources (databases, reagents, publications, teaching materials and tools), use collaboration tools and be part of the online community. (Colen, Foster et al. 2014).

2.4 Concepts of the theoretical framework

This chapter introduces briefly those theories and concepts which were used in building the theoretical framework, presented in Chapter 4.2.

2.4.1 Technology Acceptance Model

The Technology Acceptance Model (TAM) provides a detailed explanation what factors determine the acceptance and actual use of VREs. The original TAM is a framework which helps understanding why people accept or reject computer systems. Davis et al. argue that people's intentions of using a system depend on their attitude, which is a combination of different factors, like subjective norms, perceived usefulness and perceived ease of use. From all these factors, in their experimental setup, perceived usefulness influenced people's intentions the most (Davis, Bagozzi et al. 1989). However this model was introduced a while ago, it can be used to understand the acceptance of other computer-based systems, like internet usage (Porter and Donthu 2006), internet-banking (Lai and Li 2005) or to car navigation systems (Park, Kim et al. 2015).

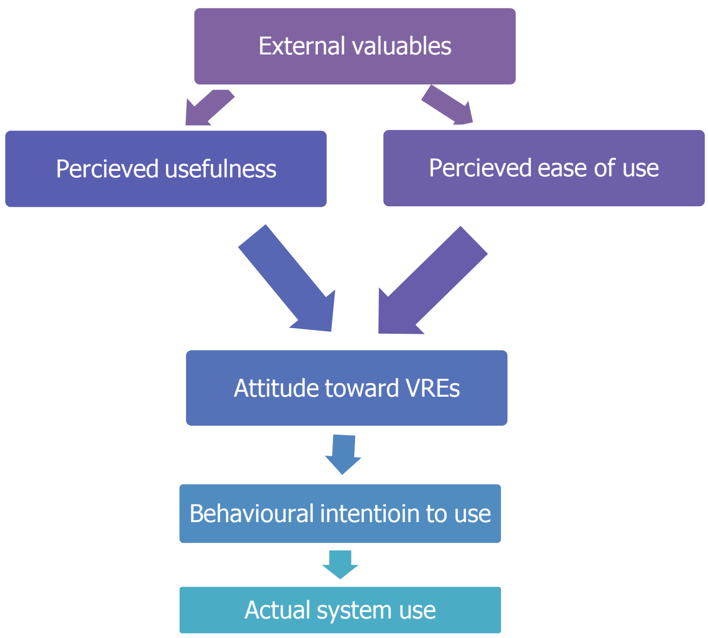

Figure 6. Technology Acceptance Modell (based on Davis et al 1989)

2.4.2 Network theory

Social networks join people who know each other. Such a network can be visualized as a group of points, representing the individuals, and the lines connecting them, representing their interaction with acquaintances. Communities at work, school, sports clubs can be treated as social networks. Extensive research has been performed to study these kinds of networks, to investigate their behaviour, the way networks help the spread of information or diseases. Scientific work is a social process. Social interactions are present in all steps of the research life cycle, from the development of new ideas to the execution of the research tasks. Therefore, the social network theory can be applied to scientific collaborations. The networks of collaborations can contain weak and strong interactions between the members (Newman 2001). Networks containing relatively many weak ties are thought to be good for scientific collaborations because they allow smooth information flow between different groups when information means simple knowledge while strong ties are important in the transmission of complex problems and coordination of complex tasks (Walsh and Maloney 2007).

Group interaction can be manifested not only in the interaction of the individual members, but the behaviour of the group as a whole can be significant. Group cohesion facilitates communication and encourages cooperation. Therefore, Walsh and Maloney argue that collaborations with more strong ties have fewer problems in project coordination related and trust issues (Walsh and Maloney 2007).

3 Methods

I have used a mixed method during my research. To respond to the first sub-question, I was checking scientific journals and books to see what was already studied and published on the acceptance or rejection of VREs by users. Parallel with this approach I performed interviews with the authors of some the publications I have found during the literature review and with some other VRE researchers and developers. Their hands-on experience guided me to develop the questionnaire and the survey used to answer the third sub-question. Finally, I have performed semi-structured interviews with eight investigators and performed a survey, which was filled out by 54 researchers. Based on these I could determine the factors that play a crucial role in influencing life science researchers' perceptions about Virtual Research Environments.

3.1 Literature search

To find those studies which describe the possible or tested factors that determine scientists' attitude towards VREs, I have performed an extensive literature search through diverse scientific domains, covering Information and Communication Technology, Computer-mediated Communication, Research Policy, Technology Management and Sociology. I have read scientific articles, project thesiss and books as well. I have searched for literature related to scientific collaborations, e-infrastructure, virtual research environments and all related topics. The topics and the used keywords are listed in **Table 1**.

Topic	Used keywords
Virtual Research Environment	Virtual research environments, collaboratories, virtual communities, types of collaborations, types of collaboratories
Collaboration	Scientific collaboration
Virtual communities	Virtual communities, CMC
Communication practices	Communication in scientific collaborations, VRE communication tools, ICT tools, scientific publication,
Social networking sites	Social media scientists
Data management	Saving data, sharing data in scientific collaborations
Project management	Project management in scientific collaborations

Table 1. Keywords used for literature search

Sampling method: During the literature search I have used Google`s academic search engine (scholar.google.com) because compared to the web of science, google scholar led to more significant results in the initial search processes. Because it is not considered as reliable as Web of Science, I have chosen articles from the list of publications found by Google Scholar by using a list of criteria. Only those articles were selected, which appeared in a peer-reviewed journal with an impact factor and the findings are underlined by research. The articles chosen by the help of these criteria were ranked based on their content and relevance to the topic. I have downloaded the articles using the TU Delft credentials. Books, which cover the topic of e-science, e-research and VREs were also employed in the literature review, they were found via Google Scholar as well. The criteria for good books were the followings: published by an academic publisher and written after 2000. The thesiss used this thesis were suggested and written by the VRE developers I have talked to. After reading the articles, thesiss and the books, they were fed into an EndNote file, and this program was used to make references in this thesis.

3.2 Interviews with VRE researchers and developers

To get an overview of what do those people think about the key factors that determine scientists' attitude towards VREs, who actually work with and develop these infrastructures, I have performed open interviews with professionals who have expertise in this area. As the majority of the VRE literature I have found was written a while ago, I expected to get an up-to-date overview of the current VRE landscape.

Sample method: I have started with interviewing Annamaria Carusi and Thorsten Reimer, who have written a collaborative landscape study about the JISC-funded VREs (Carusi and Reimer 2010). They have suggested me to contact several other people who have expertise in this topic. This is how I have contacted Matthew Dovey, who was the VRE project manager at JISC a few years ago. He suggested getting touch with SURF in the Netherlands, which also had VRE-related projects in the past. From SURF, Jan Bot answered my email, and he provided me a list of experts to contact. From this list, I have interviewed several people. In the meantime, I was also talking to people from TU Delft, and finally, I was checking which HORIZON2020 projects were successfully receiving fund on the e-infrastructures section and had spoken to researchers who are located in the Netherlands.

Interviews: In the interviews with VRE developers I have performed open interviews: I have used open questions about their work, their experiences with and opinion about VREs. The interviews were performed either face-to-face or via Skype, and they were recorded with the permission of the interviewees. First I introduced myself, the research and the interview setup, and then asked questions. Later on, I transcribed the interviews manually with the help of the InqScribe software. The transcribed interviews can be found in the separate Appendices document. The transcribed texts were then subjects to qualitative analysis. At this step, the interviews were labelled with the name of the experts.

For the analysis, I used InVivo 11 software to link the predefined set of codes to the interviews, which were based on the theoretical framework. The coding trees can be found in **Chapter 6.1.** The interviews were analysed by retrieving context attached to given codes which belonged to three major categories of the theoretical framework.

3.3 Interviews with scientists

Semi-structured interviews were performed with eight life science researchers, four members of the EpiPredict consortium and four active users of the we-NMR platform. EpiPredict consortium was chosen as a case study because via Nadine Bongaerts, who is one of the founders of Biotecture, I become involved in the science outreach part of the consortium. We-NMR came into the picture when I have contacted Alexandre Bonvine as a VRE developer. The two groups are differing from each other. On the one hand, EpiPredict is an international collaboration performing research on breast cancer-related epigenetics. It is an actual collaboration, where all the members know each other, have made a project plan together and already started to work on their joint goals. On the other hand, we-NMR is a virtual research environment, which provides data analysis services for structural biological studies. The members of this platform do not necessarily know each other; do not form an active group, but they all use this e-science tool.

Sampling: I have contacted fourteen members of the EpiPredict consortium, received (both positive and negative) answers from eight people, and could arrange four interviews. I have received eleven contact details from Alexandre Bonvin, the coordinator of the we-NMR project. Six of them answered my email, and I could arrange four interviews.

The design of the interview questions: I have used the theoretical framework and the results of the interviews performed with VRE developers to design the interview questions. The detailed description of the design process can be found in the introduction of Chapter 6.1.

Interviews: Semi-structured interviews were chosen to reduce error due to the variability of circumstances. The interviews were performed either face-to-face or via Skype and were recorded with the permission of the interviewees. First I introduced myself, the research and the interview setup, and then asked the same set of questions. Most of the questions were closed, but there were some open-ended questions as well to gain personal opinions. The interview questions and the rationales behind them can be found in **Chapter 6.1** (page 65).The recorded questioned were transcribed manually with the help of the InqScribe software. The transcribed interviews can be found in the separate Appendices document. The transcribed texts were then subjects to qualitative analysis. At this step, the interviews were labelled with a number to provide the promised anonymity to the interviewees.

For the analysis, I used InVivo 11 software to link the predefined set of codes to the interviews, which were based on the theoretical framework. The coding tree can be found in Chapter 6.1. One coded interview is presented in the separate Appendices document. The interviews were analysed by retrieving context attached to given codes which belonged to three major categories of the theoretical framework.

3.4 Survey

The survey was designed to cover most of the issues that turned out relevant in the semi-structured interviews made with life science researchers to see how general these findings were. The questionnaire is a mixture of open and closed questions, with the abundance of closed questions. I had chosen to use open questions when I did want to limit the answers by giving choices. For example, I found the answers given for the question *What do you think about collaboration*? very informative in the interviews. In most cases the interviewees gave insights into the motivations why they collaborate, what kind of collaborations do they form or participate in, and so on. That is why I left this question open in the survey as well. I also asked scientists' opinion about social media and VREs to gain an overview of their attitude about these topics. The detailed design process and the questions can be found in Chapter 7 (Table 8. The survey questions on page 80), and the answers given by the respondents can be found in the Appendices.

Pre-test: The survey was tested on two formal colleagues of mine. During the pre-test, there were no problems, so I felt the questionnaire as it was.

Getting respondents: I have designed an online questionnaire by using a survey-making tool, SurveyMonkey (https://www.surveymonkey.com). This tool provides the opportunity to distribute the survey via mobile, the web and social media to reach a wider target audience. I have put the link to the online questionnaire in several LinkedIn groups, targeting life science professionals, posted it on facebook, twitter and our scientific blog. I have also sent out emails to my LinkedIn contacts, asking them to distribute it to their acquaintances. I have closed the survey after a month, in total 76 respondents filled it out. This sample size represents 10% margin of error, at 90% confidence level and 0.5 standard deviations.

Analysis: All the data related to the questionnaire were downloaded from the SurveyMonkey.com homepage in an Excel-compatible form. In total 76 respondents completed the survey. 18 were discarded due to incompleteness. In 8 cases the respondents did not participate in any collaboration, and the survey was designed in such a way that at this point the respondents were directed to the end of the survey. In 10 cases some problems occurred in the server and the scientists could not finish the questionnaire. 4 datasets filled out by professionals from social sciences, or humanities were used, because of unequal distribution compared to life sciences professionals. The Excel table containing the data from in total 54 respondents was opened in the IBM SPSS Statistics version 22 software, and the textual answers were transformed into numeric variables. For the closed questions, where there were limited answer possibilities, this method was straightforward. Some of the open questions were turned into a Likert scale based on the manual coding of the given answers. As in the case of Question 16 (Please tell me your opinion about social media), answers were grouped into five categories based on the attitude they represent (sceptical, slightly negative, neutral/mixed feelings, OK, positive). These transformed answers were then transformed into numerical variables.

Quantitative analysis was performed by using the IBM SPSS Statistics version 22 software. Chi-square tests were carried out on the datasets, except the answers to the multiple choice questions and the transformed answers given to the open questions. I have chosen 0.1 as significance level, to see all potential differences between the different groups. For the multiple choice questions, multiple response analysis was performed in crosstabs and the frequencies were compared to see if there are any tendencies, but these results were not subject to statistical analysis.

4 Literature review

Focusing on answering sub-question 1 (What key factors did previous studies find that determine life science researchers' attitude towards VREs?) I have performed an extensive literature study. The details of the methodology are described in Chapter 3.1 (page 32). In the first subchapter, I present those studies that discuss the potential and proved key factors that affect the acceptance or rejection of virtual research environments. Subchapter 4.2 introduces the theoretical framework build from these studies and relevant concepts introduced in Chapter 2.4 (page 29).

4.1 Collaboration readiness

Although scientific users are thought to be early adopters and promoters of ICT technologies (Teif 2013), the initial attempts of collaboratories were not always successful (Bos, Zimmerman et al. 2007). Interestingly, it turned out that most of the early VRE projects had almost the same learning curve, which means that the best practices from the earlier projects were not helping the development of later ones. These projects were experiential, and the previous examples did not provide learnable skills, the developers had to have the hands-on experience to see whether their design can work on not (van der Vaart 2010).

Several studies are focusing on the determination of factors that play important roles in determining scientists' digital choices. The major elements can be categorized differently. Dutton groups them into six categories: economic, legal and public political, geographical, technical, institutional and finally social and ethical (Dutton and Jeffreys 2010).

Users of virtual research environments have to change the way they otherwise work, communicate or solve tasks and they have to adapt to the virtual environment. Olson et al. have summarized their experiences with different scientific collaborations in a paper in which they claim that not every community is ready for using collaboration assisting ICT tools. They have articulated three dimensions that play a crucial role in accepting and using these tools: collaboration readiness, collaboration infrastructure readiness and collaboration technology readiness (Olson, Teasley et al. 2002). As these three dimensions more or less overlap with the three dimensions I was focusing on (aspects at personal, group and technology level), I have listed the found factors under Olson`s categories.

4.1.1 Collaboration infrastructure readiness

At the technology level, the infrastructure is the basis of the collaboration assisting tools. The infrastructure has to be adequate and function properly in the given virtual environment to have a working collaboration tool (Olson, Teasley et al. 2002). Above the technical details of the technology itself, there are some aspects of the infrastructure which are crucial in influencing the attitude of scientists towards VREs. As VREs are online collaboration tools, to understand these aspects behind the utilization of these tools, Shackel's (1991) basic factoring could be used. Shackel disseminated three main attributes of the infrastructure that must balance the costs of the tool (including financial costs, as well as social and organisational consequences) to lead to acceptance. The first attribute, **likeability** means to what extent the users feel the tool suitable. The second attribute, **utility** refers to the match between user needs and product functionality (Shackel 1991). It is a general problem that the actual service implemented in the VREs is not what the users needed (Carusi and Reimer 2010). Some of the infrastructures were designed in a way that potential users were not included in the whole process. Even when the users were asked about their *needs and requirements*, these requests were usually assessed superficially. One of the reasons for this phenomenon is that most of the researchers communicate and articulate their needs poorly in software development language. The user-centred design would be an ideal solution, but that requires an entirely new perspective from the developer`s side (Olson, Teasley et al. 2002). The third attribute of infrastructure is **usability** which refers to users' ability to utilise the functionality in practice. Usability tests are measures how successfully users can reach some particular goals by using the tool. Ease of use (including flexibility and learnability), effectivity and finally satisfaction (success rate in reaching a goal) determine the actual user's perception of the usability of the product (Shackel 1991).

Large-scale integration and interoperability play key roles in influencing usability by increasing the effectivity of the VRE and by providing the opportunity to use a wide variety of tools within one single-sign-on environment. Most of the projects have developed brand new applications for their VREs to provide specific functions, but the majorities of these systems do not integrate the tools and software used by researchers. Most of the scientists have their favourite DNA or protein database, software to categorize and visualize plasmids, cell lines, publication systems or research tools. If it is not possible to integrate these ICT tools into the infrastructure, the transition from the remote communication and research practices to a whole new system will be extremely hard. Interoperability between different systems is also an important factor for the usability of VREs. If a VRE cannot provide seamless authentication and authorization, the users would turn away with a higher chance (Carusi and Reimer 2010, van der Vaart 2010).

At the introduction phase of the environment, it is strongly advised to provide strong **technical and instructional support** (Olson, Teasley et al. 2002) to enhance the learnability of the infrastructure. Facing technical problems, especially in the learning phase and the lack of training can cause great obstacle against the usage of the infrastructure. Users just turn back to their previous legacy systems (Carusi and Reimer 2010, van der Vaart 2010). In some cases, the developers use *cutting-edge technologies* in the creation of VREs, and the users do not know (so they need to spend the time to learn them), trust and, therefore, accept these still evolving technologies (Carusi and Reimer 2010).

4.1.2 Collaboration technology readiness

The first collaboration assisting tools began to appear as end products in the mid-90s. Since then, advanced tools started to spread in and across research communities. The adoption of VREs, like the adoption of other modern ICT technologies, has a normal progression. People are sometimes reluctant to use new tools if they already have a tool or a way of working for one particular task (Allan 2009). Moreover, scientists who do not even use simpler technologies, for example, to archive data, can have difficulties in adopting advanced technologies. Bartling and Friesike have grouped scientists into two categories based on how they **manage and communicate information**. Scientists, who belong to the **1.0 way**, track the version changes manually, send out emails with presentations and the manually tracked documents to the collaborating partners to work on and use data transfer protocols to transfer large files. On the contrary, scientists who use the **2.0 or the cloud way**, use cloud-based tools that synchronize folders within the working group and create a version history by itself and use collaborative authoring tools to work simultaneously on the same documents (Bartling and Friesike 2014).

Based on similar observations, Olson et al. advise mapping the technology readiness of the scientists before introducing an advanced level of infrastructure (Olson, Teasley et al. 2002). If the majority of users are unfamiliar with most of the tools available in the VRE, it worth to start with using a limited number of basic tools, which are easy to use and indeed needed for the given community, and extend the list of applications later on (SURF 2011).

ICT tools that are utilized in a Virtual Research Environment can be grouped into different categories. These categories are listed in **Table 2.**

Communication tools	Emails, Chat, Instant messaging Video tools, Voice over IP Manuscript submission and review systems Digital libraries Wikis, blogs
Social networking tools	Profile, CV, publications, networking, discussion fora
Project management tools	Including calendar tools
Workflow management tools	MyExperiment
Data management tools	Shared repositories, Electronic notebook
Software used for research	Software for data generation, analysis, visualization

Table 2. ICT tools used in a VRE (based on Allan 2009)

The use of these tools differs from scientist to scientist, and the attitude of scientists towards these different ICT tools can be essential in determining their attitude towards the acceptance of the whole infrastructure. For that reason, I searched for literature focusing on the practices of using each of these tools one by one, starting with communication tools.

4.1.2.1 Communication tools

Collaborations cannot operate without communication. Emails and video conferencing are highly accepted ICT tools in an international communication setting, but most of the scientists prefer to meet **face-to-face**. Informal two-way communication is essential to ensure not only the flow of information but also to create a shared culture between the collaborating partners.

> *Physical proximity helps scientists avoid or minimize many of the problems that arise in the process of conducting research - meeting partners, defining problems, planning projects, supervising co-workers and subordinates- and may influence the probability of repeated collaboration" (Kraut, Egido et al. 1988)***.**

At the beginning of collaboration research, proximity was counted as the major driving force of collaboration. Proximity was suggested to work through supporting low-cost but high-quality communication between partners. At that time, when international collaborations just started their increasing trend, it was proposed, that remote collaborations would require technologies that make communication cheap, frequent and informal enough to resemble face-to-face meetings (Kraut, Egido et al. 1988).

In virtual teams, communication plays a central role to affect performance. Teams that function in a virtual environment, rely heavily on ICT tools and this dependence on computer-based communication can cause obstacles in information exchange (Qureshi, Min et al. 2005). As the interdependence between the collaborating partners increases, the demand for more interactive and intensive forms of remote communication is also increasing. Several studies have shown that procedures that keep the **information sharing and problem-solving** smoother, result in higher productivity (Walsh and Maloney 2007).

Social networking sites offer the possibility of informal and interactive ways of communication, and in the business world, they have already become channels used for collaborating and improving innovation by engaging communities, enhancing two-way communication between different parties and incubate innovative ideas (Moore and Neely 2011). As Virtual Research Environments contain elements that could be categories as Social Networking Sites, the acceptance of and the use of VREs extremely depends on how scientists view social media and social networking sites.

4.1.2.2 Social Networking Sites

Virtual Research Environments have a part, which can be categorized as a Social Networking Site (SNS). These sites require that the members create a profile to define their presence in the web-based system. The members then add connections to other members, creating a list of associations. Finally, users can navigate through such associations to access a wider network (Orchard, Fullwood et al. 2014). Besides the general sites, there are more and more closed social networking sites specialized for scholars, like LabSpaces, Academia, Mendeley, Research Gate, Surgytec, Sciatble and so on (Van Eperen and Marincola 2011).

These academic networking sites were shown in various studies to facilitate communication between scholars and enhance scientific collaboration (Jordan 2014). But besides the strictly academic part of the social networking sites, there are other beneficial effects these sites can offer to scientists. With the wide-spread use of social media, blogs and wikis, scientists can explain their scientist results, and the societal - ethical implication of their findings directly to a wider audience without the filter of the conventional media. Due the interactivity of the web2.0 and social media sites, there is a possibility to start a discussion, so these ICT tools are found to be ideal **for public engagement**. The traditional **scholarly communication** is also changing thanks to social media. The majority of scientific papers, conference summaries, and poster abstracts are published mostly digitally and spread or publicized by science blogs and social media sites (Bartling and Friesike 2014).

Studies performed to investigate the **acceptance of web2.0** (including general and professional social media sites**) among scientists** show widespread scepticism. Many scientists see Facebook and Twitter unsuitable for professional use, and that social media use just consumes too much time. Van Eperen et al (2011) have found out only 13% of the asked scientists used web2.0 frequently for scientific purposes, 45% of them used it occasionally, and 39% did not use web2.0 at all (Van Eperen and Marincola 2011). The sceptical attitude towards social media found in this study is contradicting to the latest PEW thesis (2015), which claims that the younger generation finds social media sites, like Facebook and Twitter more important for their career advancement. So it suggests age-dependent differences between scientists. This thesis also states that at least 47% of the 3748 members of the AAAS (American Association for the Advancement of Science) were using social media to talk about science or read about scientific results (27% does this regularly and 20% only rarely) and 24% of them write blogs about scientific issues (PEW 2015). This research, which was performed five years later then the Procter study showed a difference in the percentage of scientists who use these media frequently (13% in the 2011 study and 27% in the 2015 study). This difference can arise not only because the two types of research were performed at different time points, but from lots of various circumstances, as these studies are not entirely comparable. Regarding **social networking sites designed for scholars**, this picture is a bit more sophisticated. An online survey asking scientists about their awareness of SNSs showed that roughly the 88% of those who filled the questionnaire out knew Research Gate, and roughly 55% were using it. The other professional academic sites were not that well-known or generally used. Only 29% of the respondents were aware of Adacemia.edu, and only 5% of them used the site regularly. Mendeley scored in the middle, 48% knew about the site, and 8% visited the site regularly (Van Noorden 2014).

Nentwich and König (2012) categorized researchers into five groups, based on their presence and activities in academic social networks (Nentwich and König 2012). They have defined several categories of the passive academic SNS users, as can be seen in **Table 3.** Scientists, who only have a simple profile, have barely contacts and are passive, are grouped into the first class of *Me-too presence*. They claim that this group is the most frequent among scientists. In the second group, scientists are still inactive but have a more detailed profile. They use these sites more or less as a *digital calling card*. The next level of mostly inactive users is the category of passive networkers. These researchers have a detailed profile and perform some networking activity by accepting or rejecting the automatic contact suggestions or by searching for already known colleagues. At the next level comes the *active networking and communication* within the social networking site and at the top level stands the so-called Cyberentrepeneur, who performs administrator activities above the active site usage.

Me-too presence	Simple profile, Few contacts, Almost never active
Digital calling card	More detailed profile filled up with contact data, research interests and publication list, Hardly ever active
Passive networking	Users with a digital calling card searching irregularly for already know acquaintances, reacting to the SNS`s suggestions, Infrequent communication with contacts
Active networking and communication	Frequently online, Using other services of the SNS as well, Participate in forum discussions Active search for potential networking partners
Cyberentrepeneurship	Active in networking and communicating Moderator, animator or administrator in groups Feedback to site developers

Table 3. Type of researchers based on their activity on academic SNSs (based on Nentwich et al 2012)

While it is thought that social media is mostly used by the younger generation, it turned out that the use of web2.0 tools is not that age-dependent, but there are rather **gender and discipline-related** differences among scientist. Men were found to be the two-thirds of frequent users while women were a slightly in the majority of non-users. With no surprise, the most frequent users were computer scientists, where those who used web2.0 tools the least were medical and veterinary scientists (Procter, Williams et al. 2010).

Bozeman and Corley argued that scientists' scientific, technical and social knowledge and skills add up to one collective measure, the **scientific and technical human capital** (S&T human capital). It is built on all the formal education and training, social relations and network connections. They claim that collaborations are excellent possibilities to increase this S&T human capital, as the it provides possibilities for increasing the social networks as well and develop new skills (Bozeman and Corley 2004). This is a new way to look at the competencies of scientists, by adding social networks to their portfolio. This approach can be linked to the **measurement school of e-research**, which focuses on extending the impact factors system with new ways of information sharing as public outreach or public engagement (Bartling and Friesike 2014). Web2.0 and social media are based on the idea that users generate content. For overloaded scientists, it may take too much time, especially if this kind of work is not even accredited. By accrediting activities outside the academic scientific publishing, there is a chance that scientists would spend more time in engaging with communication ways and maybe with social media as well (Van Eperen and Marincola 2011).

Besides communication and social networking, project management is also crucial in scientific collaborations, and project management tools represent an essential part of VREs.

4.1.2.3 Project management tools

From a project management perspective, scientific collaborations are rather complex projects. First, the complexity is arising from the large number of partners who are potentially coming from another scientific discipline and or from another country, which causes potential misunderstanding between collaborating partners due to cultural differences and the different jargons. The second layer of complexity is caused by the uncertainty of scientific research and the high risk taken by scientific collaborations (Beukers 2011).

Beukers have interviewed project managers of life science projects and summarized their experiences. She highlighted the skills they have found as success factors in scientific collaboration. These project management issues, which are summarized in **Table 4**, appeared in several scientific collaborations. These challenges can jeopardize the joint work, and it is hard to solve these problems without an experienced project management professional.

Person-specific communication	Stakeholder management	Team motivation
Focusing on shared goals	Leadership without authority	Monitoring
Focus on progress, not content	Clear tasks and responsibilities	Planning
Time management	Personal performance	Setting priorities
Conflict management	Negotiation with partners	Influencing

Table 4. Project management success factors in scientific collaborations (based on Beukers 2011)

Although international and interdisciplinary collaborations all have to struggle with difficult project management tasks, in life sciences, especially in academia, the importance of project management is underestimated (Allan 2009, van der Vaart 2010). On the contrary, data management is considered as a crucial element.

4.1.2.4 Data management tools

We do not have much information about how scientists save, share and manage their data. Kaizer and Kuipers investigated what kind of problems scientists at TU Eindhoven face, related to data management and safety. They have concluded that scientists, in general, **do not focus on data management**. Data plans are frequently written by grant application managers, so scientists are not aware of the promised processes or even of the data management possibilities. Moreover, the costs of the data management are not planned in the budget of the research projects. Most researchers, **therefore, use the cheapest possibilities for data storage**: central and local servers if these are free, if not, then hard drives of computers and USB sticks. Data storage costs are dependent on the size of the data, but for most scientists it is **hard to predict the amount of data produced** per year (Kaizer and Kuiper 2014).

Funding agencies and scientific journals have the tendency to require the scientists to **upload their data into public repositories after publication** (Kaizer and Kuiper 2014). It has been shown that sharing the raw data in public repositories has significant positive effect on the citation of the publication, but this positive impact is not realized by scientists (Piwowar, Day et al. 2007). Two researchers have tested how often scientists skip this requirement. They have randomly checked ten publications appearing in PLoS Medicine or PLoS Clinical Trials, and they have found that none of randomly chosen ten author groups deposited their raw data in any repository. Upon request they have received one set of raw data, the others either did not respond or refused to provide access to the data their results are built on (Savage and Vickers 2009).

Basic	Disaster recovery - backup and archive	Access control - authentication and authorization	Archiving for long term use	
Initial	Data stewardship	Metadata use	Handling data privacy in place	Training and workshops
Reuse and sharing	Data curation - shared and indexed data in searchable repositories	Data Lifecycle management	Data standards – naming conventions, ontologies	High-level data interfaces
Flow-based management	Integration of workflow and data management	Platforms to handle data, workflow and applications (VRE)	Connected to global communities	

Table 5. Levels of data management (based on Kaizer and Kuipers 2014)

Based on their research among scientists, Kaizer and Kuipers have developed a category system, which is visualized in **Table 5**. This system provides help in categorizing scientists or collaborations into distinct data management levels. At the *basic level*, scientists do very simply measurements against disasters, like backing up their data and basic archiving. At this level authentication and authorization are already present in the access control, at least at the university level, and the published data are archived for longer periods. *Initial data management* begins with distinct data stewardship, defined at the beginning of each project with a focus on data privacy, use of metadata, training and workshops on this topic. *The third level of data management* is including data reuse and sharing. Data is curated in searchable data repositories, data life cycle is taken into account, standards for data managements are used in naming files and using metadata, and finally high-level data interfaces are used. The highest current level in *data management is flow-based.* Scientists at this level integrate workflow and data management, use platforms to handle these issues, and are usually connected to global communities (Kaizer and Kuiper 2014). This categorization system suggests that only those scientists can see the added values of VREs, who are at the highest level of data management.

4.1.3 Collaboration readiness

Olson et al. call this dimension as *collaboration readiness* which can be a bit misleading as their whole theory is about research groups' readiness for collaboration. In their view, collaboration readiness is the degree to which team members are motivated to work together; thus this dimension is not focusing on individuals, but on how these individuals form a group (Olson, Teasley et al. 2002).

Scientific work is a social process; collaborations are social interactions. Social interactions are present in all steps of the research life cycle, from the development of new ideas to the execution of the research tasks. This gives the basis for applying the **social network theory** to scientific collaborations. The networks of collaborations can contain weak and strong interactions between the members. On the one hand, networks containing relatively many weak ties are thought to be good for scientific collaborations because they allow smooth information flow between different groups if information means simple knowledge. On the other hand, strong ties are important in the transmission of complex problems and coordination of complex tasks (Walsh and Maloney 2007).

Group interaction can be manifested not only in the interaction of the individual members but in the behaviour of the group as a whole. **Group cohesion** facilitates communication and encourages cooperation. Therefore, Walsh and Maloney (2007) argue that collaborations with more strong ties have fewer problems in project coordination and have less trust-related issues (Walsh and Maloney 2007). Moreover, the adaptation of technology, especially in the case of the collaboration-assisting virtual environments, has a social component, which requires that all members of the given collaboration conforms to a common pattern of interaction and settled rules (Qureshi, Min et al. 2005). Communities that already existed before the introduction of the cyberinfrastructure worked better in the virtual space than those, which were formed at the time of the creation of the VRE (van der Vaart 2010). Community uptake, elaboration and maintenance of social networks are central to the sustainability of the VREs as well (Olson, Teasley et al. 2002, Carusi and Reimer 2010). If senior researchers are not involved in the active use of the infrastructure, just students, the VRE will probably not be sustainable.

A supportive local environment is also crucial; it has been shown to play a vital role in the acceptance and spread of web2.0 tools among scientists (Procter, Williams et al. 2010). Functions that are related to long-term infrastructure building do not fit into a typical academic position. People who are involved in the development of a VRE need broad and deep programming, communications, and data management skills. Although this is a very specific job, in several universities and projects Ph.D. students receive the role of developing VREs (Lawrence and Wilkins-Diehr 2012). People responsible for the management and governance of a VRE are usually scientists from a particular domain or those who understand the technology of the infrastructure (or both). In the case of a growing and developing infrastructure, a community organizer, an activist is also needed (Lawrence and Wilkins-Diehr 2012). The role of the supportive local environment (research institute) is critical in providing job possibilities for these positions.

Collaborations which are formed for a particular call to a funding agency are paid as groups. Financial issues therefore also appear on the group level. Funding, especially the prospect of **continued funding** is a significant factor in the data management decisions, as well as in the uptake of VREs (van der Vaart 2010). Dutton mentions the example of the

Gemini telescience tool in his book. The funding of this project in the UK was reduced; the membership of the project was cancelled, which led to limited access to images from the internationally connected telescopic resources by the U.K. astronomers. Several scientists left the U.K. when finally the funding of the project was restored (Dutton and Jeffreys 2010).

Interdisciplinary collaborations result not only in broader perspective, collective expertise and higher chances of gaining grants but can also cause misunderstanding and loss of trust due to the **differences in the languages and the work practices** (Carusi and Reimer 2010). There are cases, in which **legal, ethical and cultural issues** are the obstacles in using a VRE due to different users. A typical example is when fear of losing ownership of data and ideas stands in the way of sharing (Carusi and Reimer 2010). There are disciplines, in which data are kept by the researchers, and only the interpretation of data is sent out to the collaborating partners. Forcing these researchers to share all the data that they kept private before publishing could result in a negative attitude towards the infrastructure itself (van der Vaart 2010). Social, ethical and cultural norms also shape the ICT choice of academics, just as well as the **institutional policies** of the universities or research centres where the members of the collaboration are located. Intellectual property right, liability and copyright codes strictly regulate the choices scientists have to make in accepting and using a given infrastructure (Dutton and Jeffreys 2010).

Olson and his colleagues have described several components that determine the group level of adoption of VREs, like motivation to collaborate, shared ideas about collaboration, and experience in specific areas of working together (Olson, Teasley et al. 2002). Above these factors there others as well, which affect the adoption of VREs by collaborating groups, like collaboration types, different disciplines and cultural differences.

4.1.3.1 Collaboration type

Procter and colleagues state that there is a positive association between the degree of adoption of web2.0 tools and the involvement of scientists in collaborative research. Those who were working with different institutions (interinstitutional collaboration) were more likely to be frequent users of web 2.0 tools. Moreover, one of the drivers of occasional and frequent web2.0 users was to enable collaborations with tools like wikis, blogs, bookmarking services and referencing tools (Procter, Williams et al. 2010). This suggests that the acceptance of VREs could be different in various types of collaborations.

Much research has been done to classify different types of collaborations; these studies focused basically on business organisations and strategic alliances. Hogue (1993) suggested various stages of collaborations based on the purpose, the structure of decision-making, and the nature of the leadership of the given collaboration. He argued that there are six levels of collaboration: 0) coexistence (no collaboration), 1) networking, 2) cooperation, 3) coordination, 4) coalition and 5) collaboration (Hogue 1993). Most of these categories do not fit scientific collaborations, as even the most loosely formed scientific collaboration depends on shared information and the roles are clearly defined, which means that the mildest collaboration starts at stage 3 in Hogue's definition.

Based on interviews performed in a science centre, Hara (2003) suggested that there are two main categories of scientific collaborations. The most general form of scientific teamwork is **sequential interdependence**, and as it only requires awareness and complementarity, it is rather easy to establish. Scientists in this kind of partnership work on the same project, but they are responsible for their parts of the project. In the end, they provide the results of these sub-projects, and the result is finally bigger than any member could accomplish on its own. This kind of collaboration is frequent in the field of chemistry and biology (Walsh and Bayma 1996), due to the complementary expertise of the scientists. Collaborations in which participants work together throughout the whole research process, from developing ideas, through carrying out the research and solving occurring challenges, to summing up the results are called as **integrative collaborations**. This type of collaboration requires more respect and trust, and can lead to more conflicts over responsibility and contribution (Hara, Solomon et al. 2003).

Different types of collaborations have different attributes. Based on Hara and his colleague (2003), collaboration type is determined by the motivation for working together (Hara, Solomon et al. 2003).

4.1.3.2 Motivation for collaboration

Hara and Solomon (2003) found some critical factors that are important in forming a scientific collaboration, and one important factor is the reason why scientists want to work together. On one hand, **external incentives**, such as funding, publications, and prestige, are significant in sequential collaborations. These are fundamental factors, which make collaborations possible, but these are external factors. Without funding there is no research, the publication is the desired outcome of the research process, the basis of efficiency, and prestige is the measure of professional success. On the other hand, integrative collaborations require internal motivations, such as learning new technologies, solving complex problems or fun (Hara, Solomon et al. 2003).

Collaboration types, as well as the motivation for working together, are different among scientists from different disciplines.

4.1.3.3 Disciplines

Scientists from various scientific fields were shown to show different tendencies in working together. For example, physicists collaborate more than social scientists. The data they deal with during the research is also different (from text files of interview transcripts through photos of gels to raw crystallography data), so the way they share and analyse these data is also different. Scientists from various disciplines have also found to have different priorities regarding the importance of the distinct elements of the research lifecycle. For a huge percentage of researchers doing biological science, it is not important to do deal with the managements of research outputs (Allan 2009).

Based on these differences several scholars suggested, that different disciplines would adopt VREs with varying rates, due to the different cultural obstacles they face and due to the specific skills they need to learn to be able to use this infrastructure (Sonnenwald, Solomon et al. 2002, Allan 2009).

Markauskaite (2012) mapped the ICT-use of scientists from various disciplines and have found an interesting contrast between physical and biomedical disciplines. Physical scientists used more frequently computer-mediated communication tools (20% used web2.0, 11% used collaborative writing tools) and digital tools for data sharing (24%) compared to biomedical scientists (2%, 2% and 9%), although researchers from the biomedical field were more intensive involved in collaborative work (Markauskaite, Kennan et al. 2012). Another study investigating the web2.0 use of scientists found also discipline-dependent differences. They have found that the most frequent web2.0 users were among computer scientist while those who used web2.0 tools the least were in the group of medical and veterinary scientists (Procter, Williams et al. 2010).

Based on his research comparing 53 different collaborations from the field of physics, Chompalov has generated four categories of research cooperation: bureaucratic, leaderless, non-specialized, and participatory. Chompalov argues that participatory collaboration is unique to the field of **particle physics**. These collaborations belong to the less bureaucratic type of partnerships and are highly egalitarian. Moreover, decision-making is participatory and consensual, the structure is based on shared understandings rather than formal contracts, and the data is always gathered collectively. All the other three collaboration categories are valid for other disciplines, although field sciences, such as **geophysics** has more structured and tightly managed projects, while less strictly organized projects can be found in the **laboratory sciences**, such as materials science (Chompalov, Genuth et al. 2002).

4.2 Theoretical framework

Virtual Research Environments are ICT tools that are designed to assist the various aspects of scientific collaborations. Although scientific users are thought to be early adopters and promoters of ICT technologies, the initial attempts of VREs were not always successful (Bos, Zimmerman et al. 2007). To find out why and to answer the first research sub-question (What key factors did previous studies find that determine life science researchers' attitude towards VREs?) I have performed a detailed literature search through various disciplines. The key factors explained in the different thesiss; books or scientific articles were merged into a systematically built theoretical framework.

Users of the Virtual Research Environments have to change the way they otherwise work, communicate or solve tasks and they have to adapt to the virtual environment. Olson and his colleagues (2002) have highlighted three dimensions that play a crucial role in accepting and using these tools: collaboration readiness, collaboration infrastructure readiness and collaboration technology readiness (Olson, Teasley et al. 2002). As, regarding the factors that belong to these categories, the three dimensions were more or less overlapping with the three dimensions I was focusing on (factors at personal, group and technology level), I have decided to use Olson`s categories.

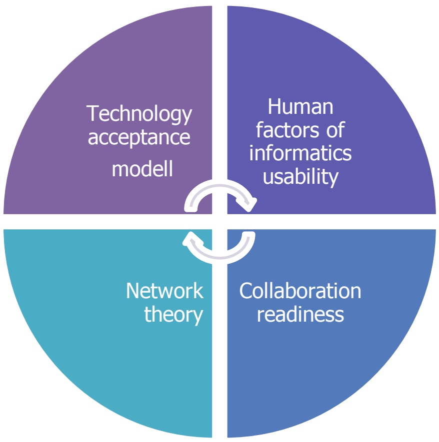

Figure 7. The building blocks of the theoretical background

Figure 7 illustrates that during the process of building the theoretical framework, I merged four different theories into one scheme. The basis of the framework is the **Technology Acceptance Model** because Virtual Research Environments are ICT tools that need to be accepted. The TAM theory was extended with Schakel's **model of human factors for informatics usability** (Shackel 1991). Furthermore, I assigned the three different factors of the Technology Acceptance Model (namely the external variables, perceived ease of use and the perceived usefulness) to the three major dimensions of Virtual Research Environments discussed by Olson et al as **Collaboration readiness**: collaboration readiness, collaboration infrastructure readiness, and collaboration technology readiness (Olson, Teasley et al. 2002). Collaborating partners function at different levels in these complex social networks: on personal level, as individual researchers, with particular motives and goals; users of the given ICT-tools and also on group level, as members of the collaboration, of a research group, of a research institute, of a scientific discipline and of a cultural community. The **network theory** is used throughout the whole thesis to provide an overview of the different levels of the collaborative network.

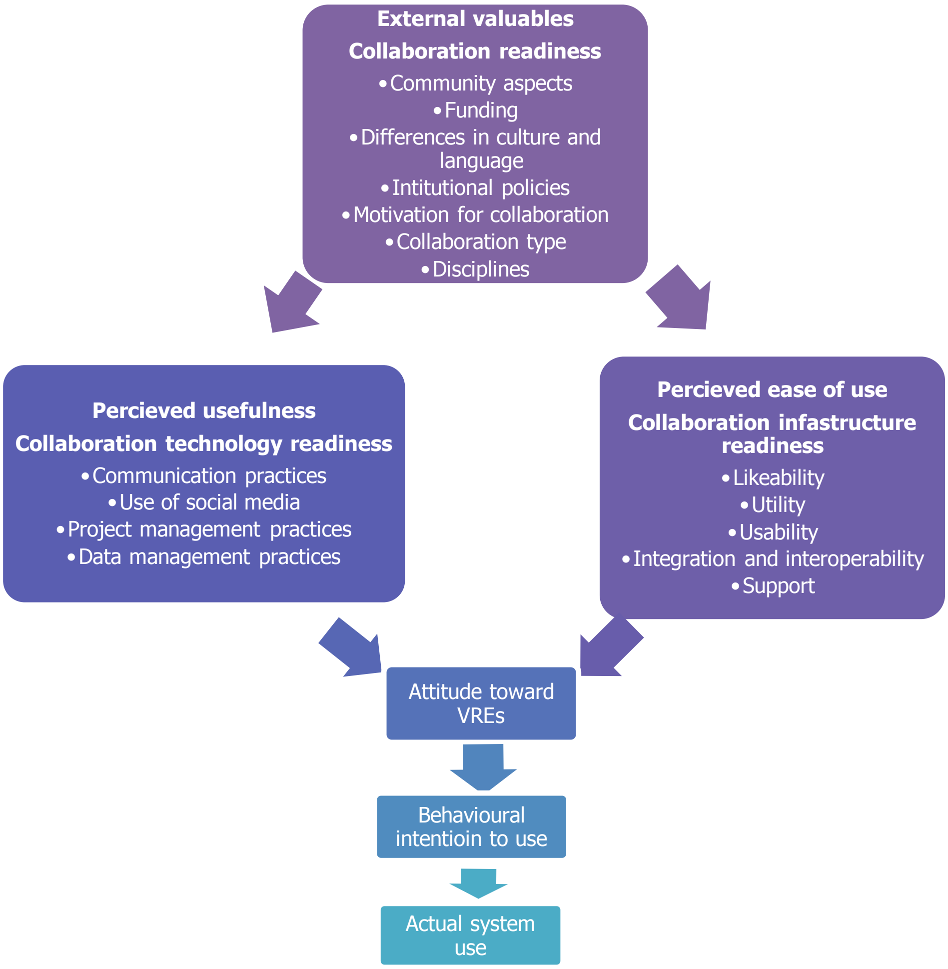

Figure 8. The Theoretical framework of virtual collaboration tool readiness

In the framework, which is visualized in **Figure 8**, I assigned collaboration readiness with external factors from the TAM, because collaboration type, disciplines or motivation for collaboration can influence the usefulness and the ease of use of VREs. Collaboration infrastructure readiness, which focuses on the technological aspects of the adoption of VREs, corresponds to the ease-of-use in the TAM, as the technical details crucially determine how users perceive the given infrastructure. I relate the last segment of the Technology Acceptance Model, the perceived usefulness to the personal level, especially to the level of collaboration technology readiness. The communication, data and project management practices and the use of social networking sites by the individual researchers determine what elements of the VRE researchers find elementary or completely useless. To successfully implement an infrastructure, developers need to be aware of on what level the potential users are on ICT-usage.

In the following parts of the thesis I use this framework to guide the presentation and the analysis of results. This framework served as a basis also for the coding tree implemented in the qualitative analysis of interviews.

5 Results of the interviews performed with VRE developers

To answer the second sub-question (What do VRE researchers and developers think about the relevance of the factors found in the literature?) I have interviewed seven people who have hands-on experiences with building or investigating virtual research environments, science gateways or e-infrastructures.

In this chapter, I present the analysis of the interviews with the help of the theoretical framework and conclude them briefly to help the understanding how these results lead to the next step in the research.

5.1 Methodology

In the interviews with VRE researchers and developers, I have performed unstructured interviews: I asked questions about their work, experiences with and opinion about VREs, and the discussion had a free flow. The interviews were performed either face-to-face or via Skype, and they were recorded with the permission of the interviewees. VRE R&D professionals were chosen based on the recommendations of researchers and other VRE developers or system providers. The profiles of the interviewees in alphabetical order:

Alexandre Bonvine is the project leader of the we-NMR Horizon2020 project. **Annamaria Carusi** is doing currently a research focusing on computational science for life sciences and for biomedical science. She is the author of the Collaborative landscape study (Carusi and Reimer 2010). **Matthew Dovey** has started to work with VRE projects at Oxford University between 2005- 2006. Then he joined JISC, where he was program director for digital infrastructure and within that context they have funded a number of VRE programmes. Now he is working in JISC on research e-Infrastructures and research technology requirements. **Rob Hooft** is the technical coordinator for the Dutch division of ELIXIR. ELIXIR is a research infrastructure for life sciences. **Sílvia Delgado Olabarriaga** is a computer scientist, and she started working in the e-infrastructure field at the UVA in a project called Virtual Laboratory for e-Science, VLE, which was a large e-Science project in the Netherlands between 2005- 2010. **Thorsten Reimer** works at the moment in Imperial College in London, where he focuses on open access to publications, research data management and data publishing and related scholarly communications and data curation activities. **Zhiming Zhao** is a senior researcher at the System and Network Engineering (SNE), at University of Amsterdam (UvA). He coordinates research and development activities in the "Data for Science" theme in the EU H2020 environmental science cluster project ENVRI Plus. This project aims to join 21 e-infrastructure into one VRE. He also leads the research tasks of research sustainability in the EU H2020 VRE4EIC project, in which they plan to build even a bigger community, not only for only environmental science; but also for life sciences, high energy physics, and so on.

The detailed methods related to the interviews can be found in Chapter 3.2 (on page 33). The quotes used in this chapter were selected to illustrate the way VRE R&D professionals think about certain topics.

5.2 Collaboration infrastructure readiness

VRE researchers and developers had a broad picture of the technical aspects of virtual research environments and the factors that affect scientists in adopting these infrastructures. But what was evident after the interviews, is that there was no consensus on what VREs exactly were. The **definition of VREs** was rather plastic, almost every person who was involved in the R&D, had a different interpretation of what VREs were: part of a research infrastructure, the user community behind the science gateway, the science gateway itself, interfaces of diiiverse tools, infrastructure used for collaborations and scientific research and so on.

> ***Matthew Dovey**: "So as you are probably aware there are now computing infrastructures, HPC (High performance computing) infrastructures like PRACE (Partnership for Advanced Computing in Europe), research data infrastructures at EU level like EUDAT, cloud resources, workflow management tools, like Taverna, which underpins MyExperiment and especially collaboration tools using asynchronous communication tools like emails, forums, and synchronous ones such as Skype, what we are using at the moment, and the work on research environment is actually how to get those things interact. It might be a web portal, might be a desktop, might be a virtual machine, etc."*

> ***Silvia Olabarriaga**: "We see science gateways as systems where everything comes together, the data, the processing, and the collaboration between people."*

5.2.1 Likeability

Although Schakel's model is focusing on three categories, likeability, utility and usability, the analysis of the interviews revealed that VRE developers did not mention likeability of the infrastructure as a critical factor.

5.2.2 Utility

The experts had different views on the utility of VREs. Some of them had experiences with a variety of projects, and were already involved in the development of the early research-assisting e-infrastructure as well, and they shared their experiences of these past projects. For example, Zhiming Zhao explained the beginning of the virtual laboratory projects in the Netherlands as a necessity to deal with the complexity of the technologies.

> ***Zhiming Zhao**: "In 1998 when I started my life in this university (UVA), we had a project called virtual laboratory. After the virtual laboratory project, we had another project, virtual laboratory for e-Science. After this was the third project, the COMMIT. All these projects were big. In the beginning, around 2000., there were lots of concepts made: problem-solving environments, virtual laboratories, and scientific workflows. One of the reasons why people made these concepts was that at that time technologies were so diverse. At that time grid technology appeared as well. When you had horizontally aggregated platforms, or infrastructures, it was difficult to use them. You needed these kinds of tools, to hide this complexity, to have the scientists do the integration. For that purpose, we needed this kind of tool. Doesn't matter how they called it."*

Matthew Dovey, who was the VRE project manager at JISC, summarized the project in few words during his interview. He explained that the ground of their VRE project was based on VLEs. When big VLEs emerged and started to be separated into modules, such as course management, assessment and so on, within those frameworks. Then they have asked the question whether they could build a similar framework for research and research tools. He said that in the second phase of the VRE project they were focusing on four large environments. One of these persisted, **MyExperiment**, which is still used by scientists to share or create workflows in a social media-like environment. Dovey described it as *the Facebook for researchers*. The third phase of the VRE project, of which not so much information is available on JISC's website, resulted in a bit of disappointment, regarding Dovey.

> ***Matthew Dovey**: "Then at the third phase we sort of felt that the world had moved on, and although there were people still using portals within certain communities, it was much more about integrating tools. JAVA portals have sort of had their days. We saw a lot of people building web apps and mobile apps, and so forth. We thought: Should we move on to this more open environment of applications work together and look at the integration and not worry about shoehorning everybody into a particular web portal?". So that is a sort of broad issue of VREs from our contexts. At the end of that process we sort moved on to inter-adaptable tools and potentially environments for sharing for collaboration: sharing data, sharing software, sharing workflows, as well as research results. I am looking at how to build collaborative systems and how to integrate collaborative systems rather than... So we are actually finding the concept of VREs a little bit of a straitjacket because it came with the obligation that you have to go to the One place. So that's where we reached. And just as we were reaching that conclusion, Europe became interested in VREs."*

Thorsten Reimer, who was working at JISC, and co-authored the collaborative VRE landscape study, explicitly mentioned, that it is not necessary to have a VRE from the researchers side. What is actually needed is a safe and reliable data management system.

> ***Thorsten Reimer**: "What academics look for, top number 1, is anything that helps saving time, reduce administrative overhead and make things easy on the system's side, having to enter information only once, to be able to use it, share what you want to share, but keeping secure and private what you want to be private. All these things might come up. And for the specialized users, the HPC community would like having access to storage which is well-connected to the HPC facility. That's something that everyone shares, obviously. But all these kinds of requirements are shared by institutions and scientists as well."*

Rob Hooft described a given project, the **TraIT** (Translational Research IT), in which CTMM (Center for Translational Molecular Medicine) has developed an infrastructure for twenty different hospitals. They delivered a standard solution for collecting, saving, analysing and archiving privacy-sensitive biomedical data, without a VRE. He mentioned this project as a success story of standardization of data management for similar projects within a given domain and as an example of the collaboration between different health care institutions **without the need for building a VRE**.

On the contrary, with these sceptical opinions about the utility of VREs, other professionals dealing with e-infrastructures were concerned that these have added values for scientists. Silvia Olabarriaga, for example, mentioned that they were building science gateways at UVA, and scientists were using the different generations of these gateways since 2010. She also commented on the sceptical attitude of other experts. She claimed that the **difficulties in the automation of data analysis processes** could stand behind this negative attitude.

> ***Silvia Olabarriaga***: *"We do not own the data analysis methods or the expertise, we just get the software which is a well-accepted open source or free for use, and we deploy it, and we make it run on the grid on some data, that's what we do. I don't care if that's automated or if it made by a person. But I see that many bioinformaticians believe in (and I guess they are right still) that in most of the time they need a person to do that, the data analysis, they cannot automate it. And then they say they cannot automate it behind the system. And then they don't invest in the system at all. So now we have a very big discussion. In the Netherlands, I think now groups are starting to realize that they need something like this, or it would be good. We are on a crossing now because we are not a service provision group, we are a research group.*

Zhiming Zhao commented on the sceptical attitude as well. He said that he understands that viewpoint, as they faced it as well in the past, but he thinks that VREs are not adopted by scientists because **some of the VREs' aspects are not perfect**, but VREs themselves are needed. VREs make scientific research faster and more efficient. But for those scientists, who are not advanced in ICT technologies, more user-friendly tools could help the adoption of VREs.

> ***Zhiming Zhao***: *"The reason of people saying VREs are not needed, it is not because VRE itself is not needed, but it is that just specific implementations of a VRE are not perfect. If you know how to program in python some specific processes, it is needed, but for data-intensive science, there are a lot of very big problems. You need better tools. You need big clouds; you need big infrastructures to process a huge amount of data. I think integrated technologies are always needed. If you disconnect from that part, your field will become smaller and smaller. And you are far away from the Nature publication. ... One guy uses python, he needs one month, the other guy uses a VRE, he needs 1 week. You can imagine who will be the first one. So from that point of view, I would disagree with those statements. But of course, I perfectly understand why they say that."*

Moreover, Zhiming Zhao suggests a potential solution to the problem of utility: to engage the user community in the development of VREs, to start a **cooperative co-design**. They bring together developers and the actual users, the customers, as he called the scientists, and let them understand what the problem is. This cooperative co-design is also handy when the user community **change**s their **requirements.** Based on Zhiming Zhao, scientists often change their minds. But if scientists receive a standard, fixed solution, they might not need that when it gets ready. But if they could take part in the development, or make their own environment with the tools they need, customize the VRE to the way they can use it, then the change of adoption would be higher.

5.2.3 Usability

The problems VRE developers were highlighting in terms of usability were more or less covering the factors that were mentioned in the literature review. Matthew Dovey mentioned **interoperability** and the integration of infrastructure to the already used environments.

> ***Matthew Dovey:*** *"I am sort of in the position that it would be interesting to see how to integrate access to new resources and tools into the existing tools rather than try to force them (researchers) into a new environment. So if they've got life sciences VRE that they are happy with, then they can integrate within that, so for example if your workflow revolves around matlab, then the ways of integrating, using matlab as user interface into data resources or computer resources rather than forcing them to a new system might be appropriate."*

Rob Hooft was expressing the importance of **standardization** of infrastructures in the adoption process. He mentioned that scientists, who have many infrastructures for various projects, would not start to use a tool, and upload the data there if that tool is not interoperable with the infrastructure employed in another project.

> ***Rob Hooft****: "I have been in numerous projects, and each of them had their own desires. And I think that is one thing that you will encounter with any scientist participating in these projects that these are not their only projects. So if every project will ask them to use their own VRE, then it's just not handy. They will need to put all of their data in multiple research environments and kept them alive. This is probably one reason why people are reluctant to start using a system that prescribed by one of the projects."*

As Zhiming Zhao sees this question, **standardization** is the most crucial factor from all the technical challenges VRE face at the moment. He said that with standardization of semantics, data and service, a lot of issues in the VRE field could be solved: interoperability, reliability, extendibility. Zhiming Zhao claimed that the reason why the early collaborator projects were unsuccessful was that the technology they were built on changed, and the environments simply could not evolve with time.

Thorsten Reimer claimed that the adoption of VREs depends on the cost-benefit ratio, on which he understands **effectivity and learnability.**

From the other side, Silvia Olabarriaga claims that the service providers (she mentions the Dutch situation) are too general, the information what they provide and their services are not **easy-to-use**, and that is one reason for rejection. She mentioned Skype, as an example, the use of which is rather self-explanatory.

> ***Silvia Olabarriaga****: "If you, as a researcher, go to the searchSARA or e-Science centre, they are very spread, all the information is spread, the knowledge is spread. And for a medical scientist is simply a no-go. It's not completely true, sometimes you see some guys at the hospital, who have a bioinformatics background, and are actually capable of using Linux machines, and you know, filling forms with requirements in terms of CPU, etc, but this is the level of infrastructure that we have today offered for researchers, and it is generic."*

*"It has to be **intuitive and simple**, as it is to use Skype or whatever. But if you see at the same time there are things you can do with Skype, and there are other things you cannot do with Skype, and people just accept this."*

5.2.4 Technical and instructional support

Two experts agreed on that besides scientists; experts should be included in the VREs or research infrastructures, who can give advice on particular questions. Rob Hooft mentioned that in ELIXIR they provide an extensive list of specialists to the researchers, who can help in questions regarding data storage possibilities or workflow system choice. Annamaria Carusi mentioned librarians, who could be involved as archiving specialists.

5.3 Collaboration technology readiness

Almost all the VRE R&D professionals looked at VREs as a collection of tools, which cover data management, project management and communication.

5.3.1 Communication

VRE developers were not considering communication practices as factors that can possibly affect scientists' adoption of VREs. Only Zhiming Zhao mentioned that, from the technical perspective, the main challenge in the communication inside a VRE is now the **engagement of the community members**, the information exchange is not a big issue anymore.

> ***Zhiming Zhao***: *"The communication is only in terms of information exchange, it is not a big issue of course. But to help the people, the main goal of communication is now engagement. How to engage people? Not just sending an email. Let them feel that they are part of a real community, and then they are willing to contribute. That is the difficulty."*

5.3.2 Project management practices

Project management is an integral part of collaboration support, but the opinions of different experts were again contradictory. On the one hand, Silvia Olabarriaga was concerned that **scientists needed project management tools**:

> ***Silvia Olabarriaga***: *"indeed the project management is a very interesting point; I see that we lack the expertise in the tooling to do it in a right way. And we are doing project and time management the whole time, as scientists. And we don't have tools for that, or we just don't want to buy them."*

On the other hand, Rob Hooft saw project management as a service, which should be provided to scientists, and they should not be forced to use project management tools.

> ***Rob Hooft***: *"Of course, you need project management for any Horizon 2020 call, but that should be located in an office, and as soon as the office, that does the project management, starts to bother all the scientists telling them that they have to use their project management software, that's not going to work. They should not ask scientists to do their work, or to make their work easier. I think those people, and that includes me, need to make sure that we are servants for the scientists."*

Thorsten Reimer had an interesting remark regarding project management. In his opinion, the **independent scientific mindset** is an obstacle in using standardized project management tools in academia.

5.3.3 Data management practices

All researchers and developers agreed that data management systems were crucial for scientific research and for collaborations. Some had distinct views on the relationship of the data management services and the VREs themselves. Rob Hooft for example thought that VREs would only be a small part of the whole data management plan that covers the data access methods, which is defined by the analysis methods planned to be used on the data, but also covers the data archiving, or **stewardship**, which is about saving data for the time when the projects are finished. Silvia Olabarriaga saw the relationship of VREs and data repositories the other way around. She mentioned that data repositories, remote data storage places are the basis for developing data analysis software or tools, which are part of a science gateway or a virtual research environment.

Regarding the data management practices of scientists, Thorsten Reimer highlighted **the importance of institutional policies** on this issue.

> ***Thorsten Reimer***: *"If institutions were to trying promoting the quality of the data that we generated then you would see more people looking for an environment to store their data better. But institutions do not really have the interest to do that."*

Rob Hooft mentioned that although not all Dutch institutions have institutional storage to provide long-term archiving of research data, but they are planning to have or already working on building one. There are universities where already Ph.D. candidates have to have a data management plan, and university medical centres are preparing guidelines for scientists regarding data safety and confidentiality. He also highlighted that it is important for scientists to see the added values of data stewardship. He mentioned that it would be beneficial for scientists to keep track on the experimental conditions to know what went wrong (or well) in certain circumstances. It would also save time and effort in writing up the projects.

> ***Rob Hooft***: *"I'm always focusing on the fact that it's not only for others that you are keeping the data, but good data stewardship will also make sure that you actually, as a scientist, in your project, know what you did with your data, and that you don't get the problems that almost everyone has seen. Now I get the results what I want, what did I do differently from last week? What parameters did I change? Or hey, it's not working anymore, what did change? And if you have your providence in order, and you thought about your data management and you know the history of your data treatment, then you can figure that out. So that is part of the data management plan as well. And this is not only needed for others in the future in order to reuse your data but also for yourself to be able to write it up in a much easier fashion. So it saves you a lot of time at the end as well.*

Regarding data management and uploading data after publication into open repositories, Zhiming Zhao has suggested **widening the citation system**, to cite the owner of the data as well, to serve as a reward system, to motivate uploading data to repositories.

5.4 Collaboration readiness

The group level of factors which affect the adoption of VREs was present in all the interviews performed with experts. Besides the user communities, other important factors were highlighted, funding, differences in languages and culture and between disciplines.

5.4.1 Community aspects

Based on the experiences of the JISC project, Matthew Dovey drew the attention to the point that those VREs showed to be sustainable, which had a community before the implementation of the environment.

> ***Matthew Dovey****: "The successful VREs were built around communities which already existed. MyExperiment worked because they were building it when we started funding. We actually funded more work on it. The ones that failed were the ones that came together because there was a funding call; they tend to disappear quite quickly when the funding went off. And I think that's true in Europe that the most successful VREs or platforms are coming out the SPIRE projects, particularly when the SPIRE has been developed around the community that already exist, rather than trying to develop a new community. "*

Rob Hooft has articulated an interesting point regarding this topic. He said that when a **forum grows too big,** and the chance of communicating with a stranger is high, then people can have difficulties with communicating freely with each other.

Zhiming Zhao mentioned that a VRE or research infrastructure has **different communities**, communities of developers, users, trainers, educators and so on, and these communities have different needs regarding the use of the infrastructure. The role of community support is to satisfy these various needs.

5.4.2 Continued funding

The perspective of continued funding is crucial in the uptake of VREs. As Matthew Dovey commented on the European VRE projects, continued funding is also essential in the sustainability of these infrastructures. These projects which were funded by the EU received grants for a given period of time, which does not provide the perspective of future sustainability. One solution to this problem is that the financing of these VREs become part of the **standard institutional budget**, suggested by Annamaria Carusi. The institutional budget does not *per se* means that the universities pay the costs of the maintenance. In France, the research institutions are paid by the national research infrastructure, but there are platforms which became commercial, and they are paid by the customers or users.

> ***Annamaria Carusi****: "It has to become part of the standard budgeting of the institutions. That is either commercial, that often happens, that these platforms are commercial, or it has to be an educational institution, as part of the national research infrastructure, like in France for example. It can't be just a project for people in an EU project, building a VRE and hoping that the community will like it and take it up, because the community is permanent, and in three-years-time it will be gone or changed. That's why they don't want to engage with it at the first place. So that's why, if you seriously consider any of these projects, you have to have institutional backing."*

Another possibility is to use standardized protocols in VREs, so when the project is finished, the raw data, the metadata with everything else can be transported into another infrastructure without trouble.

> ***Zhiming Zhao****: "Community support, for many infrastructures in the early years, was not highlighted, because many infrastructures existed as projects. When the project was finished, the funding ended, and the community was defined as a temporal term inside the project. In FP7 and the Horizon2020 projects, sustainability becomes a keyword. So we are talking about how am I going to store the data, how am I going to keep the communities, and how to make everything sustainable after the project. But of course, it is difficult. There is money. If there is no funding, you need to do this carefully, because everybody is bounded to certain resources. There are certain ways. In the projects, we usually highlight standardization, because when you standardize everything, your tools, your services are formulated in the way, which is more reusable for the new people to adopt, and the community can be maintained. And especially when these people get into new projects, their new tools can also be applied to certain standards, so their previous results can remain. So standardization is one of the keywords there."*

But as a final success, Zhiming Zhao sees the future of VREs regarding finance issues, is that **SMEs, companies or the public invest in the budget** of VREs, invest in the developments, and that could result in the continuous growing of VREs in terms of software and community as well. In Zhiming Zhao's perspective, there was no distinction between the small-scale, collaboration-specific VREs and the large-scale, for example, EU-wide VREs from the technical side. Any VRE should have access control which can be fine-tuned to sub-groups and individual members, so in principle, a huge VRE can be used to small collaborations as well.

5.4.3 Differences in the languages, work practices and cultural issues

More than half of the experts highlighted **differences in the languages** as an obstacle to using one universal tool. While most of the VRE developers meant the actual differences between languages, for example even between British English and America English, Zhiming Zhao drew the attention to another aspect of using different languages – namely using **contradictory definitions in various scientific disciplines**. Above the non-overlapping jargons, different disciplines have different work practices as well, like using different standards for using metadata, which is not interchangeable.

The opinion of VRE developers was also contradictory in terms of the **role of cultural differences** in the adoption of VREs. Thorsten Reimer found, that it can be country-specific, how scientists think about science policy issues. This can be traced back to the legal and policy framework of a given country.

> ***Thorsten Reimer:*** *"The culture definitely plays a role: the culture of the individual institution, the country, but also the policy and legal framework in which these things are happening. I think the UK is probably at the forefront of pushing academics to publish everything open, publish data and also look after managing data. That is definitely different what I see in Germany for example. You see these differences, and just coming to the Imperial College a year ago it is also easy to see how the culture here is different; the institution is to embrace things more culturally. The employer is mostly technical, and science type of organization and the academics here are more open to systems solutions than I have seen in other universities. They have more humanities people here. The academics here are sort of automating things and make systems talk to each other. And this helps them to save time. I have seen this less so in humanities organizations in the UK. People are coming to you with these main questions, like "Is that right kind of environment that we want to be involved in - in terms of authorship or the idea of research". People say here: "Sure, if we can make it happen, and it saves me an hour, just do it". And that is a strong cultural difference just experienced here.*

While Thorsten Reimer was concerned about the role of cultural differences in adopting e-infrastructures, Matthew Dovey claimed that the principal source of difference is arising from the disciplines. Cultural issues, in his opinion, play the only role in the effectiveness of collaborations.

5.4.4 Disciplines

Experts highlighted in many cases that there are differences in the uptake of VREs between disciplines. The major distinction was made between the great scientific domains, like humanities and social sciences versus life sciences, but some experts mentioned differences between life science disciplines as well. One famous example experts liked to mention is particle physics.

> ***Matthew Dovey**: "It is probably true, that some disciplines historically collaborate, particle physics, for example, have to collaborate because of the cost and size of the particle accelerator. Each research institute can't have their own CERN. They are almost forced to collaborate. In the arts and humanities, there is much more competition of single or small-group-researchers. And there might just not be a need to collaborate. The other aspect is, again, the nature of the research. So mathematicians, for example, may not need a lot of collaboration, although there are some interesting examples of crowd-sourcing to find an answer to a mathematical question. But they not necessarily need a lot of computational collaborative tools. In let's say genome synthesis or some chemical experiments, where most of the research happen in a sort of similar type of workflows, then there is a lot of scope in sharing similar tools and workflow management.*

Going further, Matthew Dovey mentioned that it is possible that VREs are not needed or not useable for every discipline. On the contrary, Rob Hooft was arguing against making distinctions between disciplines in an infrastructure. A universal, world-wide research infrastructure or data repository should be created to be used by all kinds of scientists. These e-infrastructures should not be built specifically for one particular user community.

5.5 Conclusions

To find out what VRE researchers and developers consider as relevant from the key factors determining the attitude of life science researchers towards VREs found in the literature, I have interviewed seven people who have hands-on experiences with building or investigating VREs, science gateways or e-infrastructures.

What became apparent after all these interviews is that there was no consensus on what VREs exactly were. The **definition of VREs** was rather plastic, almost every person who was involved in the VRE R&D, had a different interpretation: part of a research infrastructure, the user community behind the science gateway, the science gateway itself, a predefined unchangeable monstrous infrastructure. All R&D professionals agreed that there was a need for integrative, interoperable infrastructure, which should help scientists in saving and sharing data. If that is what we mean under VREs, then VREs are necessary for scientists.

The **likeability** aspect of collaboration technology readiness was not mentioned by the interviewees. VRE R&D professionals had opposing views about the general **utility** of VREs. One group of professional claimed that VREs are not needed by scientists, or only distinct disciplines could utilize VRE. There were several ideas about the causes of this attitude: the third phase of the JISC VRE project resulted in a bit of disappointment; until now no good or not perfect VREs were developed; and so on. The uncertainty and plastic definition of VRE can also be the cause of this bipolar attitude in the utility of VREs. One developer has mentioned a potential solution to the problem of utility: to engage the user community in the development of VREs, to start a cooperative co-design. This idea is in line with the user-centred design suggested by Olson and his colleagues (Olson, Teasley et al. 2002). The problems experts were highlighting in terms of **usability** were the more or less covering the factors that were mentioned in the literature review (Carusi and Reimer 2010, van der Vaart 2010). The factors that appeared during the discussion were the following: ease of use, interoperability and the integration of infrastructure to the already used environments, the importance of standardization of infrastructures in the adoption process and in solving other technical challenges, effectivity and finally learnability. Regarding the technical aspects of VREs, several experts agreed on that besides scientists, experts should also be included in the VREs or research infrastructures, who can give advice on particular questions.

Regarding **communication**, in the VRE developers' perspective, the main challenge is not the information exchange from the R&D side, but the engagement of the community members. **Project management** is an integral part of collaboration support, but the opinions of different experts were again contradictory in this issue. One developer was concerned that scientists needed project management tools while another saw project management as a service, which should be provided to scientists. A third VRE expert thought that the independent scientific mindset is an obstacle in using standardized project management tools in academia. All VRE R&D professionals agreed that **data management systems** were crucial for scientific research and for collaborations. Thorsten Reimer even added that the adoption of VREs is dependent on the adoption of data management systems. This idea is similar to what which Kaizer and Kuiper suggested, that advanced data management is required to see the added values of VREs (Kaizer and Kuiper 2014). Regarding data management and uploading data after publication into open repositories, Zhiming Zhao has suggested widening the citation system, to serve as a reward system, to motivate researchers uploading their data to data repositories. This view is in line with what the measurement school of the e-research stands for (Bartling and Friesike 2014). Silvia Olabarriaga added to this picture, that data repositories, remote data storage places are the basis for data analysis software or tools, which can be part of a Science Portal or a Virtual Research Environment, so if data repositories are the first to be built, these can be used to build a VRE upon them.

The group level of factors which affect the adoption of VREs was present in all the interviews. At least half of the experts mentioned that a VRE or research infrastructure has different **communities**, like the community of developers, users, trainers, educators and so on, and these communities have different needs regarding the use of the infrastructure. The various user communities can even redefine the VRE itself as they have different needs. The role of community support is to satisfy these diverse needs. Regarding the social aspects of VREs; Rob Hooft has mentioned one significant new factor, the size of the virtual community. VREs, like other community-based infrastructure, have their life cycle too. **Institutional support** is crucial in the sustainability of VREs. A virtual research environment usually starts as a research project demonstrating the proof-of-concept. After this stage, it is developed into a functional infrastructure, which must be maintained and sustained from the available funding sources, but this is very rarely possible. Funding agencies, in general, do not support the different stages, especially the one in which the infrastructure as products is started to be used. Based on previous examples, projects, in which the development and the productization were happening simultaneously, face so many frustrating problems, that users lost their trust in the system. By the time the infrastructure was finally ready to use, the community had no confidence and sometimes even no longer need for the tools. The perspective of **continued funding** is crucial in the uptake of VREs. This can be problematic with the currently running EU-projects as well. One solution to this problem is that the financing of these VREs become part of the standard institutional budget. This is similar to the idea suggested by other researchers (van der Vaart 2010). The institutional budget does not *per se* means that the universities pay the costs of the maintenance, national research institutions can also provide budget, VREs can go commercial, so members pay for the service, or SMEs, companies or the public invest in the budget of VREs. Another possibility is to use standardized protocols in VREs; everything can be transported into another infrastructure without trouble. In Zhiming Zhao's perspective, there is no distinction between the small-scale, collaboration-specific VREs and the large-scale, for example, EU-wide VREs from the technical side. Any VRE should have access control which can be fine-tuned to sub-groups and individual members, so in principle, a huge VRE can be used to small collaborations as well. That would also be kind of sort out the funding problem, if an e-infrastructure, which is financed by an EU-budget could be used to smaller consortia as well. Experts highlighted in many cases that there are differences in the uptake of VREs between **disciplines**. The major distinction was made between the great scientific domains, like humanities and social sciences versus life sciences, but some experts mentioned differences between life science disciplines as well. One prominent example R&D professionals liked to mention was particle physics; this comes up very often in literature too (Sonnenwald, Solomon et al. 2002, Allan 2009, Markauskaite, Kennan et al. 2012). Zhimming Zhao drew the attention to another aspect of using different languages – namely using contradictory definitions in different scientific disciplines. Above the non-overlapping jargons, different disciplines have different work practices as well, like using different standards for using metadata, which is not interchangeable.

6 Results of the interviews performed with life scientists

Eight life science researchers were interviewed to find out what the attitude of life science researchers towards VREs is and what obstacles they experience that prevent them from adopting VREs (research sub-question 3). Four of them belonged to the EpiPredict consortium, a scientific collaboration focusing on EpiGenetics research, which was formed upon a Horizon2020 call. The other four scientists were active users of the we-NMR, a VRE providing crystallography and NMR calculations for their members, so they had scientific projects related to protein structure research. I have interviewed them via Skype or personally if that was possible. I have asked the same set of questions (Table 7. Interview questions page 65), recorded the interview with their consent, transcribed the interviews and analysed their answers by using a quantitative analysis tool, inVivo11. The detailed methods can be found in Chapter 3.3 (page 33). This chapter contains the results of these interviews put into the structure of the theoretical framework (explained in Chapter 4.2) whenever possible, and at the end of the chapter the conclusion of these results.

6.1 Designing the interview questions

The qualitative attribute of this part of the research determined the design of the interview questions. That was the reason for formulating lots of open questions, which provided the possibility for the interviewees to formulate the answers with their own words, highlight those aspects of the topic, which were relevant for them without leading the response to the question (Bryman 2015). I have used different types of questions, like introducing, specifying, direct, and vignette questions (Bryman 2015), for each of which **Table 6** illustrates examples.

Question type	Example
Introducing	Please tell me about your research group! How is it organized, how many people do you work with?
Direct	Do you use social media?
Specifying	Do you use it for personal or professional purposes?
Vignette	Imagine if the EU would like to build an enormous Virtual research environment, where all the data produced by an EU-founded project would go in, and all the researchers involved in these projects would have to log in. How would you feel about that?

Table 6. Question types used in the semi-structured interviews

The questions, as it is visible in **Table 7,** were grouped into various themes: research group organisation, collaboration, communication practices, data management, and VREs. During the interviews, I modified the first questions of each topic into structuring questions to let the interviewees see how the interview proceeded. The rationale for the distinct questions was based on the theoretical framework described in Chapter 4.2 and listed next to each question in Table 7.

6.1.1 Interview questions

Themes	Questions	Rationale
Research group organisation	1. Please tell me about your research group! How is it organized, how many people do you work with? 2. What is your role?	Initial warm-up, mapping group hierarchy
Collaboration	3. What do you think about collaboration in general and its role in scientific research? 4. Is your research group a member of international collaboration? 5. Are you personally a member of international collaboration? 6. What is your role in this collaboration?	Gather information about collaboration type, motivation, role in collaboration and the scope of collaboration.
Communication practices	7. What kinds of communication means do you use in this particular collaboration? 8. How do you communicate with your colleagues in general?	Gather information about the current communication practices
	9. Are you satisfied with these communication means? 10. What would you change and why?	Mapping needs regarding communication tools
	11. How do you manage the project related issues (organizing events, selecting candidates)	Mapping project management practices
	12. Is your group engaged in any forms of science outreach activity? 13. Are you engaged in any forms of science outreach activity?	Exploring science outreach practice in different levels
	14. Do you use social media? 15. Do you use it for personal or professional purposes? 16. Are you an active or passive user? 17. Do you use social networking sites that are designed for scientists? 18. What do you think about using social networking sites in the scientific field?	Gathering information about the use and the attitude towards personal and academic social networking sites
Data management	19. How and where do you store research data? 20. Does your institution or research group have a data management policy? If yes, please explain these policies.	Requiring information about practices, institutional policies related to data management
	21. How do you search for old data? 22. Do you label your data to be searchable later on?	Gathering information about metadata use
	23. How do you share data within your research group? 24. How do you share data and data-related information in this collaboration? 25. What kind of data do you share? Raw data or analysed data? 26. Did you experience problems in collaborations arising from using different data analysis tools? 27. Do you share instruments, software or visualization tools between the collaboration partners? 28. Is there a need for that in your research? 29. Would it be useful to you to add comments to files and datasets?	Mapping data, file and tool sharing practices
VREs	30. Have you ever heard about Virtual Research Environments? If yes: What are your experiences? Would you use VREs in collaborations? If no: *A VRE is an online (and potentially offline) tool which is made to help researchers by providing an infrastructure, framework and user interfaces in collaborations* (Allan 2009). What do you think this kind of tool could be used for in your collaboration? 31. Would you consider your international collaboration as a Virtual Research Group? 32. Imagine if the EU would like to build an enormous Virtual research environment, where all the data produced by an EU-founded project would go in, and all the researchers involved in these projects would have to log in. How would you feel about that?	Gaining information about VRE-related awareness, experiences, attitude and likeability

Table 7. Interview questions

I have decided to start the interview with a warming up introductory question, which helped to get over the initial inertia. The interviewees were asked to introduce their own **research group**, which is a topic they are familiar with, and the question was open enough to let them decide which aspects to share. After this initial question, I asked one specifying question to gain further specific information about their role in the research group (community aspects, collaboration readiness). Then I decided to ask about different aspects of **collaborations**, which also belongs to the collaboration readiness dimension of the theoretical framework. The first question in this thematic group was also a general open question, asking for their ideas about the importance of collaboration in scientific research. This question was aimed to collect answers regarding the type of collaboration, and potentially to determine the motifs why scientists want to or need to collaborate. The further specifying questions ask directly about the scope of the collaboration and about the scientists' role in the collaboration. The third block of questions was about **communication practices**, which belong to the collaboration technology readiness dimension. I was aiming to map the project management and communication practices, including the use of social networking sites. I was interested how scientists think about social media in general and about academia-related SNSs. I designed question to ask directly about their practices, but also about their satisfaction, needs and attitude in terms of using the current tools. The fourth theme block was focusing on **data management**, which also belongs to the Collaboration technology readiness dimension, and I formulated questions mostly direct questions to gain concrete answers on this topic. I wanted to see on which data management level the interviewed scientists were. Finally, I asked questions regarding **VREs**. I asked whether they were aware of the term, and what kind of experiences do they have with VREs, to see what factors came up from the collaboration infrastructure readiness category.

6.1.2 Coding frame

The codes used to analyse the transcribed interviews were defined before the analysis and were based on the theoretical framework.

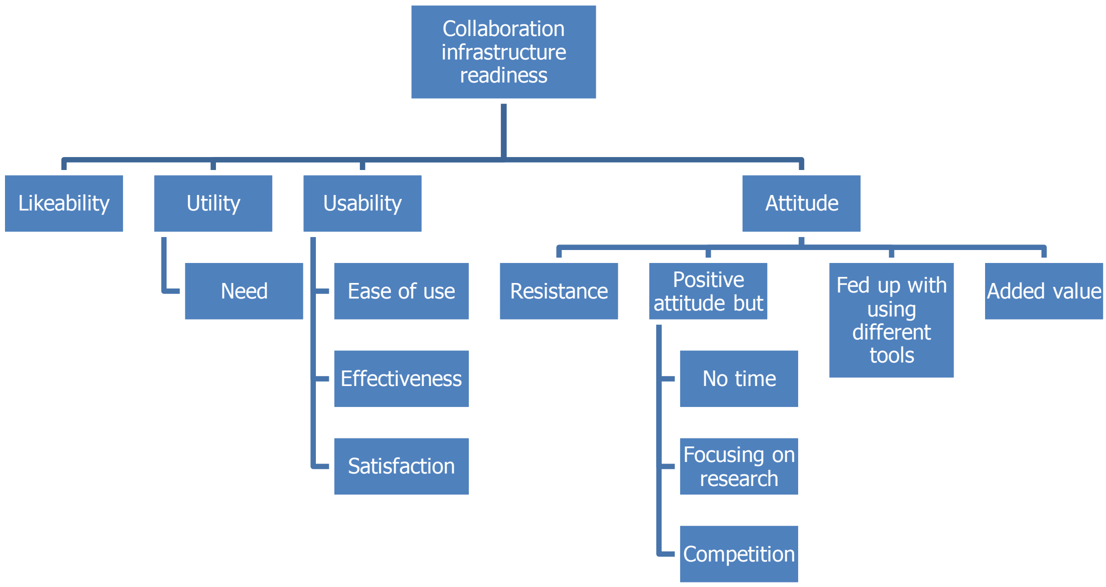

Figure 9. Coding tree for collaboration infrastructure readiness dimension

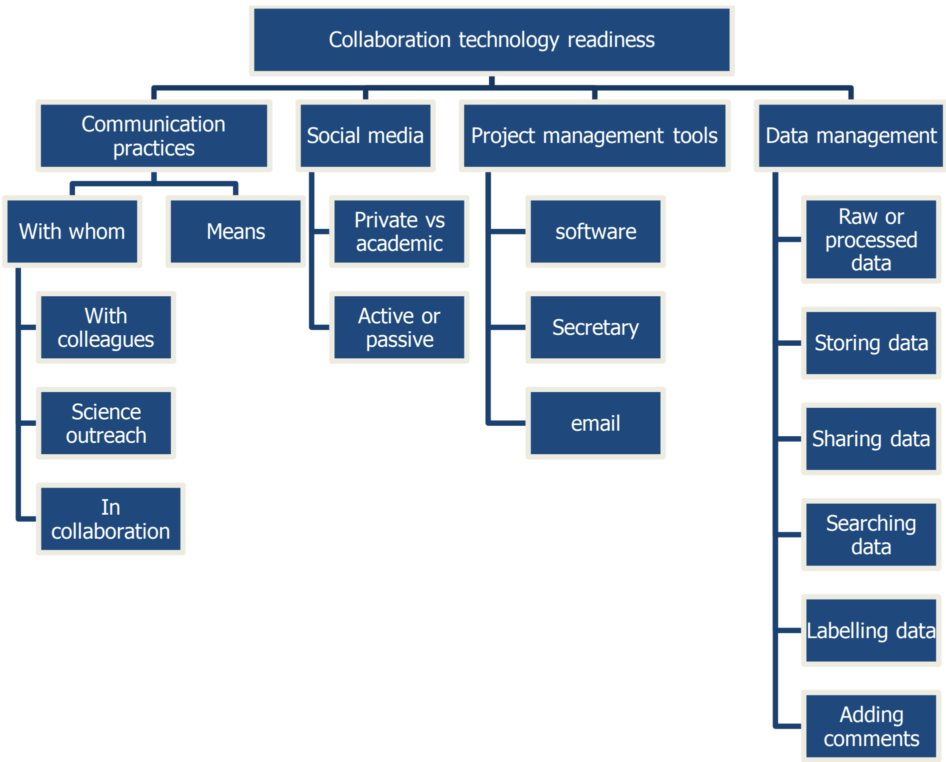

Figure 10. Coding tree for the collaboration technology readiness dimension

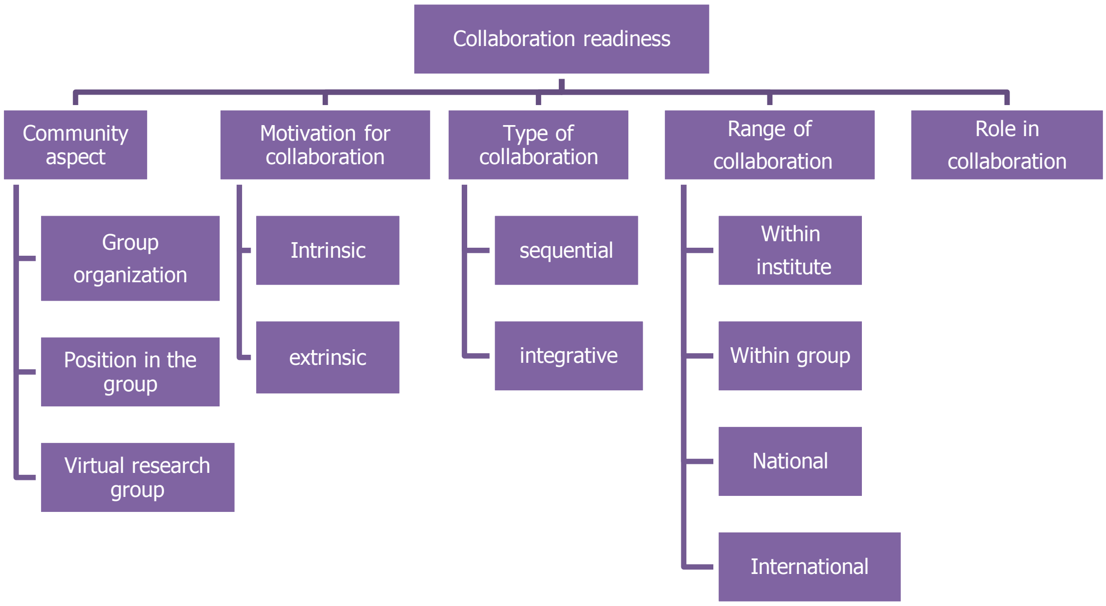

Figure 11. Coding tree for the collaboration readiness dimension

6.2 Collaboration infrastructure readiness

Half of the interviewed researchers had no idea what VREs were. Interestingly there were two scientists, who were active users of the we-NMR portal, but were not aware that it was actually a virtual community, and the portal was its virtual research environment.

Scientists who did not hear about VREs had rather a positive attitude towards this kind of infrastructure, they said that they could imagine themselves using a VRE in their collaborations, and had no objections against them. A PI from the consortium was, in general, positive about the idea of VREs, but she warned that PIs simply would not spend the time to learn a new environment.

Those researchers who had hands-on experiences with virtual research environments have highlighted their concerns about these infrastructures. There was one postdoc, who has used the we-NMR as a VRE, she said that she saw the benefits of a VRE, but probably she would not use that in full potential. One bioinformatician from the EpiPredict consortium mentioned **standardization**, **compatibility, single-sign-on system** and **utility** as crucial factors in adopting VREs:

> *#2: People get upset using this platform on this project, and another type of platform on that project. People are fed up using these competitive tools. You need to have some kind of commonality, where everyone has the same tools. Otherwise, it is big trouble. And people don't want extra passwords to log into that platform. They will resist; probably there is the laziness behind it. They will resist unless it is very helpful.*

The other bioinformatician was a PI of a research group investigating science portals, so he probably had an excellent overview of the scientific literature of VREs. He has highlighted problem like **attitude**, **learnability**, **utility**, **flexibility**, **usability and interoperability** as key factors in accepting and using VREs by life sciences researchers:

#5: I think that the concept is good. But so far it is not working, because I have never seen that clinicians or biomedical researchers are really embracing these types of environments. I think the major reason is that it takes additional efforts to learn an environment and use it, but also it does not always provide all the functionalities that people need. Or does not provide sufficient flexibility and maybe the biggest hurdle is that most of the scientists are using Windows- or Mac-based tools, which they are familiar with, and then if people are familiar with one tool, it is very difficult to have them switched to something else."

Besides the consortium-scale VREs, I asked the scientists' opinion about a possible huge VRE built by the EU, where all the data produced by EU-funded projects would go in after publication, and all the researchers would have to be part of it. All of them were positive about the idea. They have indicated several possible positive outcomes, such as this VRE could help in keeping the data accessible, make science more democratic, transparent and increase efficiency. Someone simply said that "it would help science, it would help everyone". Several researchers mentioned that this question was not that theoretical, it is already happening, there are some projects and scientific committees that work already like that, but scientists, in general, are not aware of these projects.

6.3 Collaboration technology readiness

One of the bioinformatician PIs from the EpiPredict consortium summarized why it is important to investigate the communication, social networking, data management and project management practices of the potential users of VREs. If it is not possible to **integrate** all the collaboration-related activities, which require a human-computer interaction, into the infrastructure, then the infrastructure will not be used.

> *#5: (q30) "If you want to have a virtual environment, in whatever format, it must provide data management facilities, people must have access to sufficient computing power, and they must have means for social interaction for collaboration. If one of these things are missing, then they have a problem, they have to move for one activity to the virtual environment, and then move out again to do other stuff. People don't do that. They just don't return (to the VRE)."*

6.3.1 Communication practices

The interviewees used face-to-face discussions, emails and in one case WhatsApp as communication means within their own group, and when communicating with colleagues. They used emails, Skype and face-to-face discussion when communicating with the collaborating partners. One of them were quite sceptical about computer-mediated communication for professional use, she preferred personal communications to using ICT tools for work, although she was using Facebook quite actively for personal purposes.

All of them were more or less satisfied with the communications tools they used, only two scientists wished to have more face-to-face or Skype meetings, or mentioned some problems with using Skype. One researcher mentioned that she did not like the current communication ways she had with her colleagues because these meetings were too official. It was interesting to hear that she mentioned fun, as a measure of good communication.

> *#8 (q9): "It is less fun. It's all about publications, money, Ph.D. defences; it's really these things which can be counted. We drive each other crazy. No, really, it`s not fun anymore."*

Not only were the interviewees, in general, satisfied with the used communication tools, but most of them mentioned that they also needed these tools to maintain the connection with the collaborating partners. Especially Skype, as a video conferencing tool was highlighted, that it could somehow replace face-to-face meetings:

> *#6: (q7) (About Skype) "We find it all very useful to see each other in the face, rather than just to talk on the phone. And we do that once every two weeks, so we rely heavily on that."*

6.3.2 Science outreach

Half of the interviewed scientists were actively involved in various science outreach activities. Interestingly they were mostly group leaders, or they belonged to a group, where the head of the group has encouraged lab members also to participate in these activities. Two group leaders were active in reaching out, but their group members were not. Those, who did participate in translating their scientific research to a wider audience, found science outreach a vital thing, also for scientists.

> *#6: (q13) I think it is good because we force people (group members) to think in laymen's terms. More and more grant agencies are asking larger part of the grant to be expressed in laymen's term. So it is good for them, because when they have to write a grant, they will not only focus on the science part, they will also focus on the patients' or laymen's part.*

6.3.3 Social networking sites

Scientists rarely mentioned that they liked or disliked a given ICT tool. There was one scientist, who, by the way, hated social media in general because of his bad experiences, who claimed later on that he liked LinkedIn and ResearchGate, especially the latter one because it is very well done. One postdoc was rather sceptical about social media in general; she did not even use social media designed for academic use, because she believed that the most valuable scientific discussions and decisions are made personally, not on the internet.

> *#4: (q18) I think in the academic environment, everything is done mostly by word of mouth. And a lot of connections are made in person. I don't think that people would look for a Ph.D. or postdoc on these sites, more like they know somebody who knows somebody. I think there is no real advantage in it; when you move to industry, then maybe, but in the academia, not. I am on LinkedIn because people told me that I should be.*

If scientists used social networking sites, then they were mostly passive users.

6.3.4 Project management practices

There were only two interviewees, who used an ICT-based project management tool in their lab. One of them used Microsoft Project, which enables to manage tasks, collaborate, submit time sheets and so on. The other scientist introduced QUARTZY into his lab, which is a free online lab management platform. He mentioned that his group members had difficulties in adopting this tool, and the reason for this was that it is hard to change someone's habits.

> *#6: (q11) "I see some resistance within the group. It seems like it is more work, well actually it is less work for the long term. It is always a trouble to change people's habits. "*

All the others used the same channels they used for communication, like email, Skype or asked help from a secretary in managing projects.

The effectivity of communication was mentioned by one PI, who complained that most of the meetings were frustrating. She highlighted her needs for a system, which helps to improve efficiency in running projects, without knowing what she needed was a project management tool.

> *#8: (q10) Well, what I would like to have ... is be much more effective in a sense that many meetings run without itinerary, without minutes, everybody has his personal feel what has been agreed upon, and then again because of time issues, people do not really follow up, or it ends at a lower priority, and then it would have been decided upon according to me. So to ease this kind of frustrations, it would be nice to have something, like a task list, deadline on it, and a responsible person. But many of my colleagues would not like it, they think: we are scientists, and should be creative, and should not take care of that much protocol. So I don`t think people would really favour an approach like this.*

6.3.5 Data management practices

One team PI has highlighted one critical aspect of data management practices: **fraud**, and how electronic lab notebooks could help fighting against that. She also highlighted some important aspects of archiving data for the long term.

> *#8: (q19) "As an institute we are currently exploring the electronic lab journal. That would make a real big difference in scientific integrity issues, like fraud and the grey area. It is very-very useful to have an objective, easily searchable system, but it is very expensive. You have to have lots of storage room bit-wise. People charge you for that. And the question is how long you need to save it. Now we always say that you need to save it for five years, and I am not going throw out any lab journals because you never know, maybe in another ten years, I need to see how we did things to get to this data."*

Only two of the interviewees mentioned that they have ever labelled their data, an additional one said that there was a plan to introduce a **metadata**-using system for data management. Only one scientist from the eight mentioned that they upload their data (labelled with metadata) to a data repository after publication, although this is a general custom in the structural biology discipline. Interestingly, we-NMR scientists started to think about data repositories as data saving possibilities when I asked the last question (*q 32. Imagine if the EU would like to build an enormous Virtual research environment, where all the data produced by an EU-founded project would go in, and all the researchers involved in these projects would have to log in. How would you feel about that?).*

Although six out the eight scientists did not label their data at all, half of the interviewees said that it would be needed in their discipline to **add comments to files and datasets in a collaboration setup**. One postdoc had concerns about intellectual property while a team leader had formulated his worries about false accusations. He was worrying about comments claiming falsely that the data were bad quality, and how that could be avoided.

One PI said that it was already possible with the electronic laboratory notebook system they have just introduced to add comments to data. He also mentioned that people were not used to doing that, and there would be a need for a change in the mindset to start adding comments directly to files.

> *#1: (q29) "It would probably take a long time for people to get used to the idea of doing that (labelling files and datasets), rather than just looking at the data and sending an email. But our electronic lab notebook allows people to do that. People can log in and make comments in principle. We haven't done that yet, but it is a possibility."*

Above asking about their data management practices, I have asked the interviewees about one particular extra function of VREs, namely **sharing software, analytical and visualization tools**. None of the we-NMR members had ever shared an ICT tool before while half of the EpiPredict group offered or received these kinds of tools for or from collaborating partners. Nevertheless, the ratio of those, who find sharing instruments in collaborations relevant and useful was the same in both groups (3/4).

6.4 Collaboration readiness

Based on the set of questions aiming to collect information about the research group organization, the interviewed scientists were working in hierarchically structured groups, except from one bioinformatician from the consortia, who worked in an SME. Most of them mentioned activities which served the **group cohesion**, like lab meetings, joint lunch discussions, journal clubs. In terms of the community aspect of collaborations, only half of the interviewed researchers thought about their collaborations as **virtual research groups**. Regarding **institutional policies**, it was interesting to see that none of the postdocs and Ph.D. students were aware of the institutional policies related to data management.

Most of the interviewees gave a long answer to the question: *What do you think about collaboration in general and its role in scientific research?* They all agreed that is a must to collaborate, without working together with other research groups, scientific research nowadays is impossible. Most of them claimed the specialization of science is the **main force that drives collaboration**, meaning specialization of expertise and specialization of equipment. Others mentioned that without collaboration, they could not publish their results, as scientific publishers prefer authors from different institutes.

One PI from the EpiPredict consortium explained her experiences with various **collaboration types** in detail. She had both sequential and integrative collaborations before. She mentioned that the sequential ones were more successful than those, in which scientists were more interdependent, these did not accomplish what they have planned.

> *#8: (q4)* Is your research group a member of international collaboration? *Yes, many. But the ITN (International Training Network, she means EpiPredict) is a very nice example; I think because there is really an interaction. So I have been in EU projects, where we did get together once in a while, but to be honest, everybody just did what she was doing anyway, but now from EU money. And sometimes there was a project in which you were really dependent on reagents created by another group. These projects were always problematic for me because we never got what the others foresaw to have finished in one year or two years. So there is always a problem with this kind of projects, but, at least, the intention is that there is some knowledge sharing and reagent sharing. Those have been a kind of unsuccessful projects, actually. The most successful projects have been projects where people acted independently; they got together once in a while, mainly to inform each other and maybe to get some ideas, but they are not too interdependent. That`s interesting. I think it is because I don't like that it worked out like this. But I think if this is the way, yes...*

Collaboration type showed **discipline-related differences**. Most of the structural biologists were involved in sequential collaborations while the Biomedical scientists were developing the consortium together.

6.5 Conclusions

To find out what obstacles could stop life science researchers from adopting VREs, I have performed semi-structured interviews with eight researchers from this field. Half of them researchers had no idea what VREs were, even if they were active users of the we-NMR, which is a VRE. This suggests that we-NMR is generally used for its data analysis services provided by its platform, and the VRE part of the platform is not well-known by its users.

Regarding **likeability**, all of the interviewed researchers were positive about the idea of building a huge EU-scale VRE to serve as a data repository and an SNS for scientists. They indicated that such a VRE could help keeping the data accessible, make science more democratic, transparent and increase efficiency, which is parallel to previous research results (Dutton and Jeffreys 2010). Those scientists, who were not aware of what VREs were, had a rather positive attitude this kind of infrastructure and mentioned only learnability as a potential problem for PIs. On the contrary, scientists, who had hands-on experience with VREs were aware of their potential and actual problems. The key usability-related factors in adopting VREs they mentioned were standardization, interoperability, integration, compatibility, single-sign-on system, utility, attitude, learnability and flexibility. These factors are related to **utility and usability** of VREs, and were also mentioned by VRE R&D professionals and came across during the literature search (Carusi and Reimer 2010, van der Vaart 2010).

The interviewees used face-to-face discussions, emails, WhatsApp and Skype as **communication means.** One of them was quite sceptical about computer-mediated communication for professional use; she preferred personal communications to using ICT tools for work, although she was using Facebook quite actively for personal purposes. This attitude was partially described in the study of Qureshi and his colleagues (2005), which concluded, that ICT tools can cause obstacles in information exchange in collaborations (Qureshi, Min et al. 2005). All of the interviewed scientists were more or less satisfied with the communications tools they used, only two scientists wished to have more face-to-face or Skype meetings, or mentioned some problems with using Skype. One researcher mentioned that she did not like the current communication ways she had with her colleagues because these meetings were too official. She mentioned fun, as a measure of good communication. Not only were the interviewees, in general, satisfied with the used communication tools, but most of them mentioned that they also needed these tools to maintain a connection with the collaborating partners. Especially Skype, as a video conferencing tool was highlighted, that it could somehow replace face-to-face meetings. Half of the interviewed scientists were actively involved in various **science outreach activities**. Interestingly they were mostly group leaders, or they belonged to a group, where the head of the group has encouraged lab members also to participate in these outreach activities. Two group leaders were active in reaching out, but their group members were not. It suggests that science outreach activities are heavily dependent on the group leaders' mindset. Those, who did participate in translating their scientific research to a wider audience, found science outreach a vital thing, also for scientists. The researchers rarely mentioned that they liked or disliked a given ICT tool. There was one scientist, who, by the way, hated **social media** in general because of his bad experiences, claimed later on that he liked LinkedIn and ResearchGate, especially the latter one because it is very well done. One postdoc was rather sceptical about social media in general; she did not even use social media designed for academic use, because she believed that the most important discussions and decisions in the scientific world are made personally, not on the internet. If scientists used social networking sites, then they were mostly passive users.

There were only two interviewees, who used an ICT-based **project management** tool in their lab. One of them used Microsoft Project, which enables to manage tasks, collaborate, submit time sheets and so on. The other scientists introduced QUARTZY into his lab, a free online lab management platform. He mentioned that his group members had difficulties in adopting this tool, and the reason for this is that it is hard to change someone's habits. The effectivity of communication was mentioned by one PI, who complained that most of the meetings were frustrating. She highlighted her needs for a system, which helps to improve efficiency in running projects, without knowing what she needed was a project management tool.

Only two of the interviewees mentioned that they have ever labelled their data, an additional one said that there was a plan to introduce a metadata-using system for **data management**. Only one scientist from the EpiPredict consortium mentioned that they upload their data (labelled with metadata) to a data repository after publication. Interestingly, we-NMR scientists started to think about data repositories as data saving possibilities when I asked the last question, which was related to VREs*.* One PI has highlighted one critical aspect of data management practices: fraud. She mentioned that electronic lab notebook could help fighting against the grey area in science. Although six out the eight scientists did not label their data at all, half of the interviewees said that it would be needed in their discipline to **add comments to files and datasets in a collaboration setup**. One postdoc had concerns about intellectual property while a team leader had formulated his worries about false accusations in this topic. One PI said that it was already possible with the electronic laboratory notebook system they have just introduced. He also mentioned that there would be a need for a change in the mindset to start adding comments directly to files. None of the we-NMR members had ever **shared an ICT tool** before while half of the EpiPredict group offered or received these kinds of tools for or from collaborating partners. This can be explained by the differences in collaboration type, as we-NMR scientists were all claiming that they were part of sequential collaborations, or by other factors. Nevertheless, the ratio of those, who find sharing instruments in collaborations relevant and useful was the same in both groups (3/4).

All of the interviewed scientists agreed that is a must to collaborate, without working together with other research groups, scientific research nowadays is impossible. Most of them claimed the specialization of science is the main force that **drives collaboration**, meaning specialization of expertise and specialization of equipment. Others mentioned that without collaboration, they could not publish their results, as scientific publishers prefer authors from different institutes. One PI had experienced that the sequential collaborations were more successful than those, in which scientists were more interdependent, these did not accomplish what they have planned. This in line with the findings of Hara and Salamon (2003), that sequential collaborations are much frequent and easier to manage while integrative collaboration can lead to more conflicts over responsibility and contribution (Hara, Solomon et al. 2003).

Interestingly, most of the structural biologists were involved in sequential collaborations while the biomedical scientists were developing the consortium together, showing a **discipline-specific difference**.

7 Results of the survey

To find out what obstacles prevent life science researchers from adopting VREs, I have interviewed eight researchers. To investigate how general the obstacles that were mentioned in the interviews were, I have designed an online survey. The survey was designed to cover most of the issues that turned out relevant in the semi-structured interviews. The questions can be found on the following pages (Table 8. The survey questions), and the answers given by the respondents can be found in the Appendices. I have used to design the questionnaire. This tool provides the opportunity to distribute the survey via mobile, the web and social media. I have put the link to the online questionnaire on social media to gain as much as possible respondents. I have closed the survey after a month, in total 76 respondents filled it out. 18 response sets were discarded due to incompleteness, 4 datasets because these were filled out by professionals from social sciences, or humanities and this could result in a sampling error. Finally, on the data from in total 54 respondents, quantitative analysis was performed by using SPSS. Chi-square tests were carried out on the datasets, whenever possible. I have chosen 0.1 as significance level, to see potential differences between the different groups. For the multiple choice questions, multiple response analysis was performed in crosstabs and the frequencies were compared to see if there are any tendencies. The results are presented by using the structure of the Theoretical framework.

7.1 Designing the survey

During the design of the survey questions, I was basically modifying the interview questions to serve better as an online questionnaire. The questions were grouped into the same themes with the exception of the research group organization, which I skipped, and each topic had a separate page at the survey homepage. During the design, I took into account the properties of an online survey and the requirements of the qualitative research. Therefore, I have generated more closed questions, to provide material for the quantitative analysis, although I have left some questions open. With the open questions, I aimed to collect the ideas of the participants without providing a pre-articulated answer. I have introduced questions with answers with Likert scales to measure the attitude of scientists. For the multiple choice questions, I have used the answers I have received during the interviews or have found in the literature.

The rationale behind each question was similar to the rationales used to generate the interview questions, which was based on the theoretical framework. First I introduced the research (the introduction can be found in the Appendices). After the introduction, as it can be seen in **Table 8**, I have decided to start the questionnaire with six questions (Q1 - Q6) related to **collaborations**, which belong to the of the collaboration readiness dimension of the theoretical framework. With the first question in this thematic group, I wanted to gain information about how important scientists find collaborations for scientific research. I was asking scientists to rate how strongly they support the following statement: Collaborations are important for scientific research. They could choose from five answers from strongly disagree to strongly agree. The second question was an open one because I found the open question related to collaboration in the interviews very informative. With the further questions in this thematic group, I wanted to gain information about the scope and type of their collaborations, their role in the partnership, and finally their personal motivations for working together.

The second block of questions was related to **data and project management practices** (Q7 - Q13), which belongs to the collaboration technology readiness dimension. I mostly asked single choice questions to gain specific data to see at which data management level the interviewed scientists were. There was one multiple choice question regarding the ways scientists save their data because based on the interviews, most of the people use numerous ways. I also asked a question to see how important they find saving data.

The third block of questions was about the **communication practices** (Q14 - Q21), which belong to the collaboration technology readiness dimension. I was aiming to map how scientists communicate with each other in a collaboration setup. I used a multiple choice question for this purpose, because from the interviews in turned out, that scientists use various means to communicate with each other. After this, I asked an open question (Q15) regarding their potential needs in terms of communication tools, and another one (Q16) to collect their unguided attitude towards social media. Later on, I asked more specific questions about their social networking (Q17-Q19) because I was interested how scientists think about social media in general and about academia-related SNSs. Later on, I asked questions related to their science outreach practices (Q20-Q21).

The fourth theme block (Q22 - Q26) was focusing on virtual research groups and **VREs.** I asked whether they were aware of the term of VRE, and what kind of experiences do they have with VREs, to see what factors come up from the collaboration infrastructure readiness category. I also asked a question about their attitude towards an EU-wide VRE.

The final block of questions (Q27 - Q30) was related to their personal information, like scientific domain, academic position and age.

7.1.1 Survey questions

Q1. Collaborations are important for scientific research.

- Strongly disagree
- Disagree
- Undecided
- Agree
- Strongly agree

Q2. Please describe what scientific collaboration means to you. (Open question)

Q3. Do you participate in scientific collaborations? (Multiple choice question)

- No
- Yes, I collaborate with my research group
- Yes, I collaborate within the Institute.
- Yes, I participate in national collaboration(s)
- Yes, I participate in international collaboration(s)

Q4. Why do you collaborate?

- Access to expertise
- Because it is fun
- To pool knowledge to solve a complex problem
- Access to equipment or resources
- To obtain prestige
- To improve access to funds
- To educate a student
- To learn tacit knowledge about technology.
- Because of the increased specification of science
- To enhance productivity
- To increase the chances of publication
- Other

Q5. What kind of collaborations do you participate in?

- Sequential collaboration: we work on the same project, but I am responsible only for my own pieces of the research process, contributing to the project by providing my results
- Integrative collaboration: all of us are involved in developing research problems, refining ideas, and analysing results through thesising the results.
- other

Q6. What is your role in your collaboration(s)?

- I am a participant
- I am a (scientific, administrative or financial) leader
- Other

Q7. It is very important to save the data for later use or for validation purposes.

- Strongly disagree
- Disagree
- Undecided
- Agree
- Strongly agree

Q8. How do you save research data? (Multiple choice question)

- own PC
- backing up regularly
- institutional server
- data repository after publication
- on cloud-based systems
- Other
- paper-based laboratory book
- electronic laboratory book

Q9. Do you use metadata when saving data?

- No
- No, but it would be important
- Yes, when saving my result.
- Yes, when uploading data to databases
- Other

Q10. Do you reuse data?

- No
- Only my own results
- Only the results of my research group
- Data downloaded from database
- Other

Q11. How do you share data with you collaborating partners?

- We do not share data
- By email
- By google docs
- By Dropbox
- By sharing access to server
- By sharing electronic lab book pages
- Other

Q12. Do you use workflow management tools?

- No
- No, but it would be important
- Yes

Q13. Do you use any project management tools?

- No
- No, but it would be important
- Yes

Q14. How do you communicate with your collaborating partners? (Multiple choice question)

- Face-to-face
- email
- Skype
- Social media
- Other

Q15. If you could change the communication means, what would you change? (Open question)

Q16. Please tell me your opinion about social media. (Open question)

Q17. Do you social media?

- No
- Only for personal purposes
- Only for academic purposes
- For both

Q18. Do you use any of the social networking sites designed for scientists, like Mendeley, Academia, ResearchGate?

- No
- Yes, but I am quite passive
- Yes, I use it actively

Q19. Networking is important in setting up collaborations.

- Strongly disagree
- Disagree
- Undecided
- Agree
- Strongly agree

Q20. Do you participate in science outreach activities?

- No
- My group does, but I am not involved
- Yes

Q21. If yes, what kinds of science outreach activities? (open question)

Q22. Would you consider your collaboration as a virtual research group?

- No
- Yes
- Other

Q23. Have you heard about Virtual Research Environments?

- No
- Yes, but under another name, as Collaboratories, Virtual Knowledge Centers, Scholarly Workbenches or Science Gateways
- Yes

Q24. If yes, what are your experiences with VREs? (Open question)

Q25. Virtual research environments, by integrating different data management, project management, communication, visualization and data analysis tools in a personalized way, could be very useful for scientific research.

- Strongly disagree
- Disagree
- Undecided
- Agree
- Strongly agree

Q26. If the EU would like to build an enormous Virtual Research Environment, where all the data produced by EU-founded projects would go to, and all the researchers involved in these projects would have to register, how would you feel about it? (Open question)

Q27. What is your scientific domain?

- Medical, Mathematical, Physical and Life Sciences
- Humanities
- Social Sciences

Q28. What is your position?

- Principle investigator
- Team leader
- Postdoc
- PhD student
- Other

Q29. What is your age?

- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65-74

Q30. If you would like to be kept up-to-date about the results of this research, please fill in your email address, and I send you my thesis.

Table 8. The survey questions

7.2 Collaboration infrastructure readiness

Most of the scientists have never used a VRE and have not even heard about them. From the 54 scientists filling out the survey completely, only 3 (6%) have heard about virtual research environments, while 7 (13%) were familiar with one of the other terms (Collaboratories, Virtual Knowledge Centres, Scholarly Workbenches, Science Gateways). From those, who have heard about VREs or were familiar with a synonym of VRE, only 6 used this kind of infrastructure before. From those, who mentioned that they have worked with VREs before, three shared their experiences:

"Data upload has been cumbersome and time-consuming. There are better ways to save data and organize work. "

"Mixed, most of them not working very efficiently."

"Interoperability of tools for specific functionalities is more useful, versatile and flexible than aiming to include all functionality into a one-size-fits-all solution."

Scientists had various thoughts about the usefulness of a VRE. 22 (40.7%) scientists were not sure whether VREs could be beneficial for scientific research. 25 (46.3%) of them said that they agree, and 6 (11.1%) that they strongly agree with the statement *Virtual research environments, by integrating different data management, project management, communication, visualization and data analysis tools in a personalized way, could be very useful for scientific research*. There was only one researcher, who disagreed with the statement.

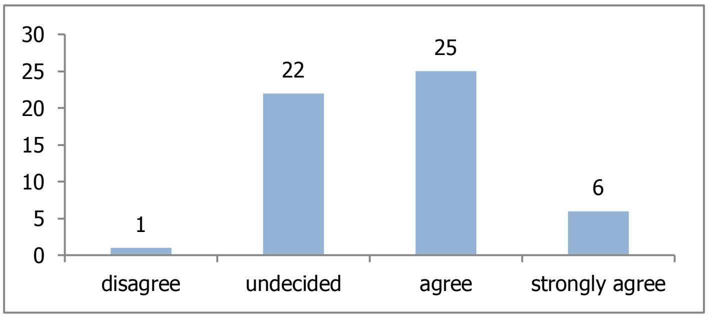

Figure 12. Scientists are not sure whether VREs could be beneficial

This is in parallel to the **highly diverse attitude towards an EU-wide VRE.** For the following question: *If the EU would like to build an enormous Virtual Research Environment, where all the data produced by EU-founded projects would go to, and all the researchers involved in these projects would have to register, how would you feel about it?* I have received a wide array of responses. By manually grouping these answers into different categories, it turned out that 8 scientists (15%) were totally sceptical about this kind project, 8 (15%) of the responses were slightly negative. 14 scientists (25%) had neutral or mixed feelings about an EU-wide VRE, whereas 8 (15%) were slightly positive, and finally 16 (30%) were absolutely positive towards this project. The reason for grouped the answers received for this open question into artificial categories was to see whether there was a relation between the attitude towards VREs and other factors, but these relations were not subject to statistical analysis, due to the artificial categorization.

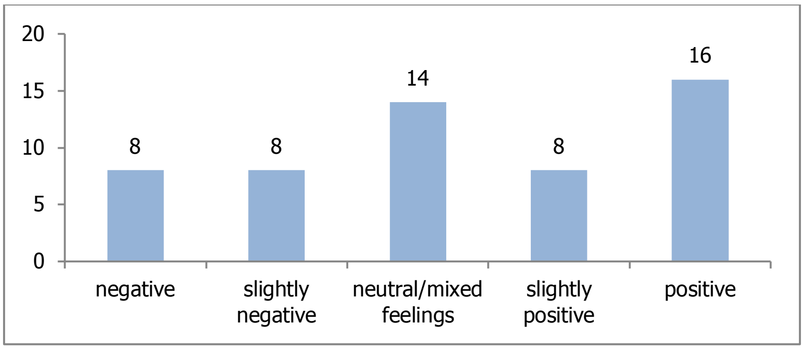

Figure 13. Life scientists' attitude towards an EU-wide VRE

The majority of scientists with a negative attitude towards the EU-wide VRE were concerned that this infrastructure would increase bureaucracy, which can result in the decrease of time spent on actual research (seven answers). Five of them were afraid that it would be too big, too general or too specific. In total, eight scientists were concerned about data safety or security, or they would not want to give out their data to the competitors or would feel uncomfortable with sharing data. Someone even mentioned that he or she wishes to control his or her own data. There was one participant concerned about legal and ethical issues related to sensitive patient data; another scientist thought that building an EU-wide VRE would be a waste of resources, a simple data repository would do better service. One researcher rationalized his or her negative attitude with the fact that it is hard to fit various types of data into one environment. Another researcher highlighted that "forcing open access publication of research funded by public money is more important". One participant was against the concept because if one action is not voluntary, then it is not good for science. One scientist had concerns that the system would not be simple enough to allow equal opportunities, access or understanding to all researchers.

From the positive side, two researchers mentioned that this kind of infrastructure could provide the place to publish negative results and detailed protocols. One was happy that the data would be freely available to the public. Two participants thought that such a VRE would be beneficial for collaborations; three thought that VREs would be useful. Furthermore, 13 researchers were simply positive about the concept, and would like to try it out.

7.3 Collaboration technology readiness

7.3.1 Communication tools

The respondents communicated by using multiple communication means. Everyone (100%) used email to contact their collaboration partners, almost everyone (94%) mentioned that he or she meets face-to-face. Skype was less frequently used, only 65% of the researchers used it, and only 1 out of 10 scientists used social media to communicate with their collaborating fellow scientists (**Figure 14**). Other communication means were also used, like a telephone.

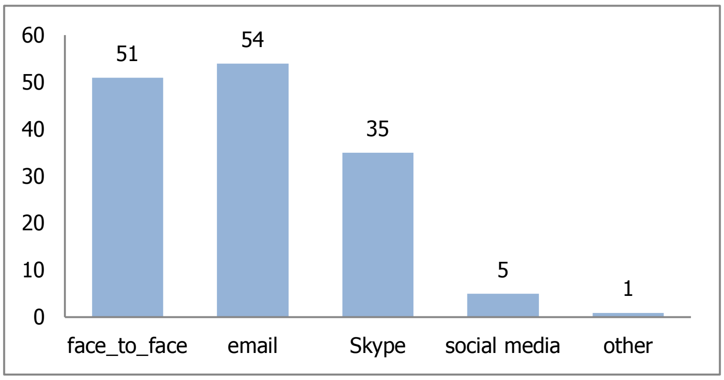

Figure 14. Communication tools used in a collaboration setting

More than half of the scientists (57%) who filled out the questionnaire **were satisfied with their current communication practices**. 19% would like to have more face-to-face meetings while others would like to change entirely different things. Some would like to have quantitative changes, like more data exchange in the communication within the collaboration, while others suggested qualitative changes, such as reducing the number of tools, to have one platform for oral and written communication. **Table 9** is summarizing these ideas.

Quantitative changes
More online video chat combined with presentation
More data exchange
Qualitative changes
Meetings should be more concise and focused
Attachments or metadata embedded into email
Communication-related to a single project could be stored in an organized way
Integration, flexibility, recording
Reduce the number of tools, i.e. have one platform for oral and written communication

Table 9. What scientists want to change regarding communication in a collaboration

7.3.2 Science outreach practices

Roughly half of the scientists (47%) in the survey were participating in science outreach activities. Interestingly, scientists bearing any kind of leader position in collaborations (p=0.078) or in the research groups (p=0.057) were more likely to participate in science outreach activities.

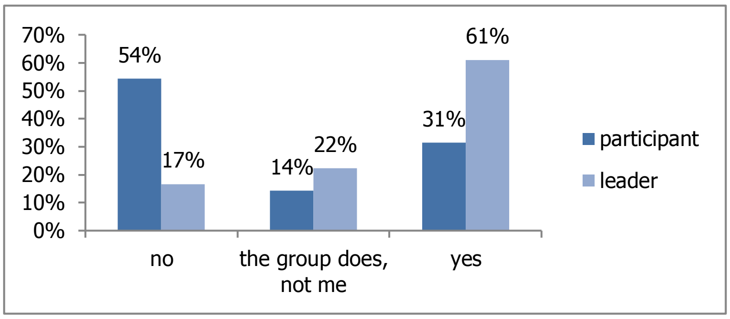

Figure 15. The result of the cross table participating in science outreach versus role in the collaboration

As science outreach, the following activities were mentioned: writing science blogs, popular science papers, managing of a social media web page open to the general public, organizing workshops, meetings, conferences for the public, Science Café's, researchers' night, student workshops, Open Days, school visits, away days for students and teachers, science week and improving public knowledge of GMO issues. Some scientists mentioned that they participated in activities such as monitoring, looking for academic partners, reference customers, supervising high schools students, dinner parties or just some casual meetings in a bar.

7.3.3 Social Networking Sites

As it can be seen in **Figure 16**, while the majority of scientists agreed (44%) or strongly agreed (48%) with the statement that *Networking is important in setting up collaborations (Q19)*, and only 7% was not sure whether this was important or not, scientists had diverse opinions about social media and its role in the academic world.

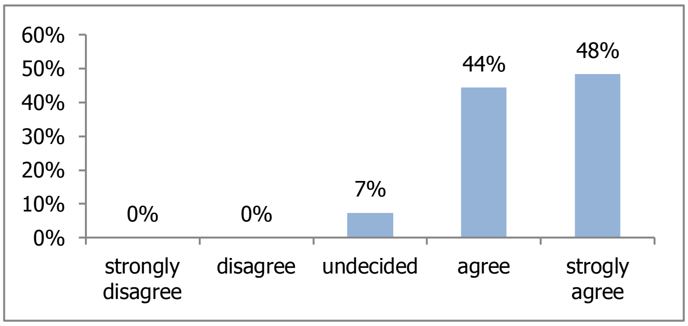

Figure 16. Scientists find networking important in setting up collaborations

In total, 18 scientists wrote down that social media was or could be useful, with or without additional concerns. The most common concern about social media was that it is only suitable for personal purposes (12 answers). A few people mentioned additional concerns about privacy, copyright and data safety (3 people), or that social media can be dangerous (3 people), or especially harmful for science (1 person), that the facts are not checked (1 person), and that it consumes too much time (1 person). Those scientists who mentioned positive aspects of social media were writing that it is useful for public engagement, education or science outreach (5 people), for networking (5 people), for funding and collaboration (3 people), for finding and disseminating information (2 people), to follow research progress (2 people) or to build a public profile (1 person). One scientist mentioned that not scientists, but specialized professionals should deal with social media.

By manually grouping these answers into different categories, it turned out that 11% percent has formulated strongly negative feelings like "overrated, too much noise and low-quality info", "useless for scientific exchange" or "the way it is used now is harmful to science". 20% percent expressed slightly negative feelings, when asked about social media and 20% percent were neutral, or had mixed (positive and negative) feeling. In total, 48% had positive or strongly positive attitude towards social media. This artificial categorization just served to see the relative amount of positive and negative opinions and to see whether there was any relation between scientists' attitude towards social media and other factors.

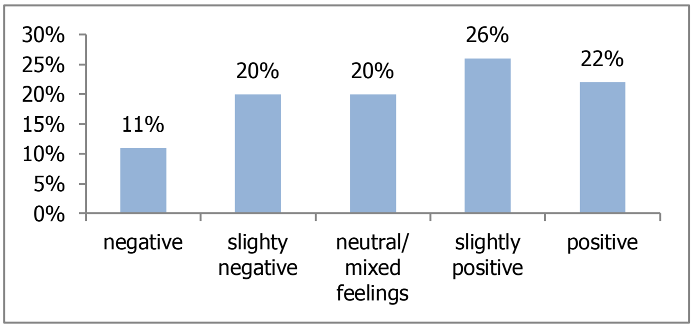

Figure 17. The attitude of life science researchers towards social media

The attitude towards social media showed an interesting age-distribution. As illustrated in **Figure 18**, the youngest age group was rather very positive or had mixed feelings towards social media. By the increase of age, more negative thought about social media sites appeared, while the last age group was relatively positive (but it was represented only by one scientist, so it is not statistically relevant).

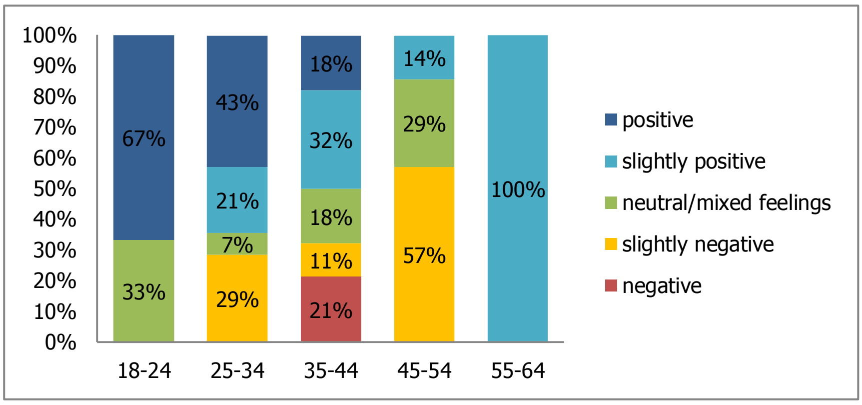

Figure 18. The age-distribution of the attitude towards social media

Figure 19 gives an overview about how many scientists used social media for personal and professional purposes. 11% of the researchers said that they did not use social media at all. These scientists gave responses to the question, *Please tell me your opinion about social media* ranging from the very sceptical to neutral, but none of them were positive about social networking sites. The **amount of scientists who used social media only for personal purposes** was huge (**46%).** Interestingly enough, the amount of these scientists was comparable with those, who were **using social media for academic purposes (43%).**

The majority of these scientists were using social media for both personal and academic purposes, but there were 2 researchers, who used these sites only for academic reasons. For the further investigation, these two groups (only academic use and using social media for both purposes) were treated as one group of academic use of social media.

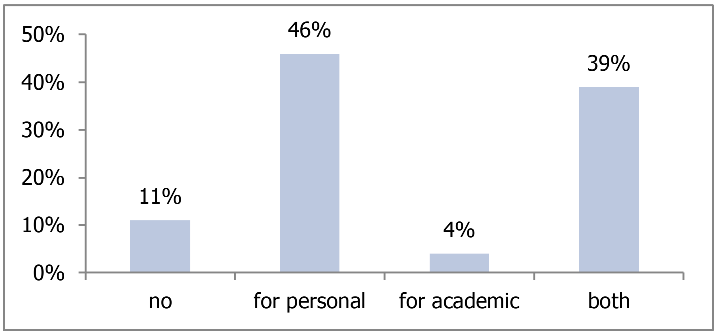

Figure 19. The majority of scientists use social media for personal purposes

Those scientists, who used social media for academic purposes were significantly more positive about social media in general (30%) than those, who only used these sites for personal purposes (20%) or even did not use it (0%) (data not shown).

88% of the respondents used academia-related SNSs, of which only 17% were active users, the rest were just using these sites passively. This percentage is much higher than the percentage of those scientists who said that they were using social media for academic purposes (43%). This discrepancy is arising from the fact, that half of the scientists who claimed that they did not use social media at all were actually admitting using social networking sites designed for a scientist, just like 88% of those, who said that they were using social media only for personal purposes, as it is visible in **Figure 20.** This figure also shows that he majority of academic SNS users were rather passive users. The highest ratio of active users was found in the group in which scientists said that they used social media for professional purposes as well (22%). Interestingly, 4% of those scientists, who said that they used social media for professional purposes, used general social media not specialized for scholars.

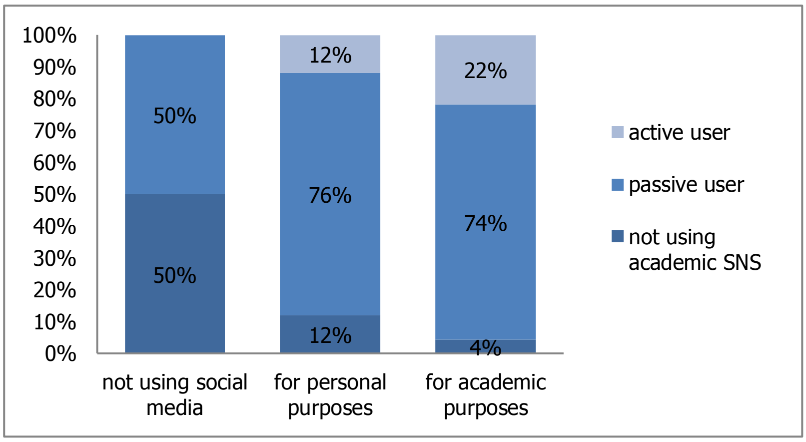

Figure 20. Half of those scientists who said they did not use social media used SNSs designed for academia

There was a significant (p=0.055) correlation between the position of scientists in collaboration and the active use of academic social networks. Scientists with a leader position in collaboration were more likely to be active users (26%) than participants (12%), while only 5% of collaboration leaders did not use any kind of academic SNSs, compared to the participants, where this number was 15%.

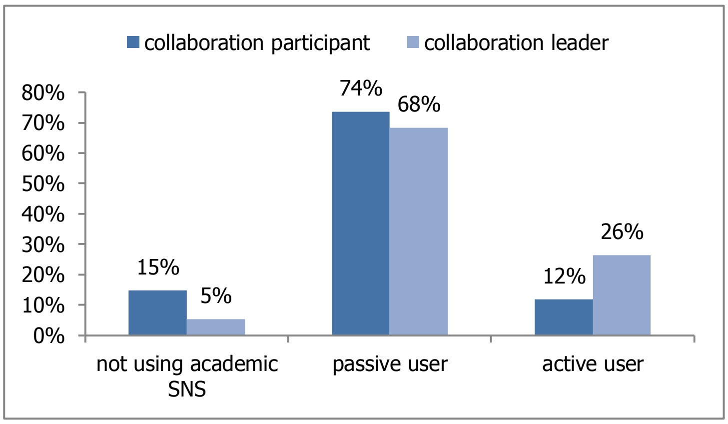

Figure 21. Collaboration leaders tend to be more active academic SNS users than participants

The results suggest a correlation between social media use and the attitude towards VREs. While in the group of scientists, who claim that they do not use social media at all, there is no one with a slightly positive or absolutely positive attitude towards VREs; those, who use social media for private or for academic purposes have more diverse opinion regarding VREs, ranging from negative to positive. 40% of the scientists using social media for personal purposes thought slightly positively or positively about VREs, while this was true for the 69% researchers using social media for professional uses.

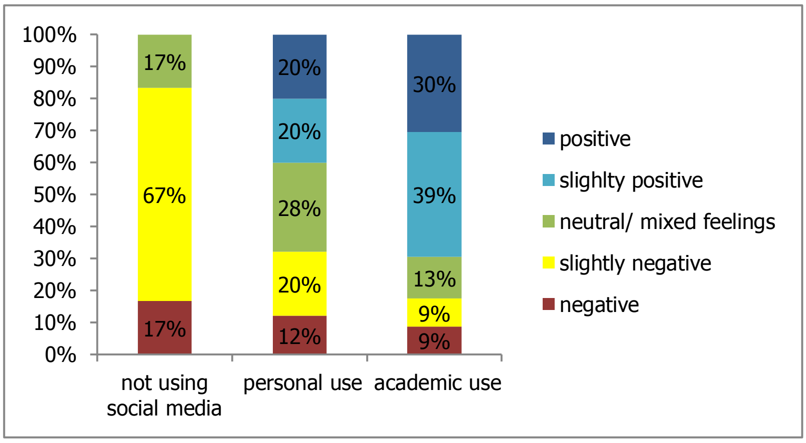

Figure 22. Scientists using social media for personal or academic reasons had more positive attitude towards VREs

7.3.4 Project management tools

Although the two tools provide solutions to entirely different problems, the distribution of scientists using project management and workflow management tools look almost the same, as it is visible in **Figure 23**. Roughly 1 out of ten researchers used any kinds of project (6 scientists, 11%) or workflow (5 scientists, 9%) management tools, and roughly 1 out of 5 (22% and 20%) did not use these tools but thought that it would be important to use them.

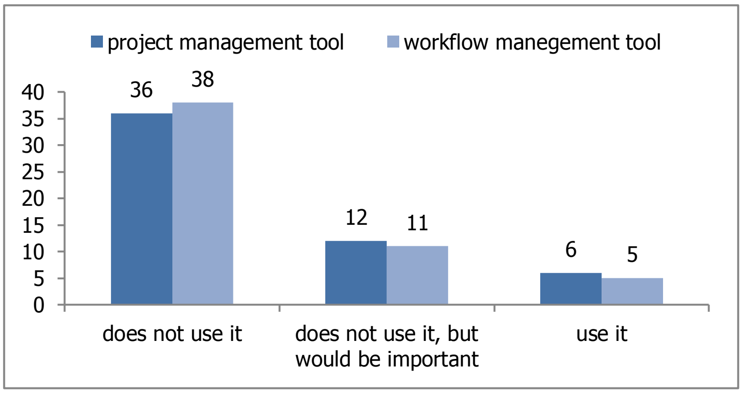

Figure 23. Only one out of ten scientists use project management of workflow management tools (presented in number of cases)

There was a significant difference between how collaboration leaders and participants thought about project management tools. As it can be seen in **Figure 24**, while the vast majority (91%) of collaboration participants did not use any kind of project management tools, 75% of them also did not find it relevant. Scientists who took any type of leader position in collaborations, used project management tools more often (16% versus 9% in the participant group), and also half of those, who did not use these tools found it rather important (42% of the collaboration leaders).

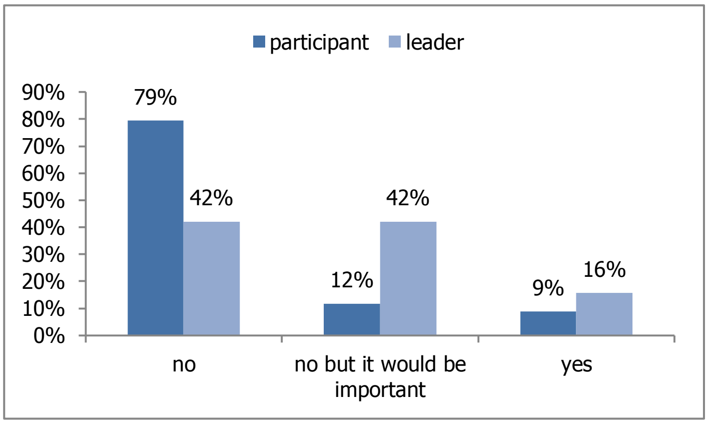

Figure 24. Collaboration leaders find it more important to use project manager tools

Those scientists, who did not use project management tools at all, as it can be seen in **Figure 25**, were mainly naming extrinsic motivations as reasons for collaborating (prestige, publication and funds). Interestingly, a huge percentage of these scientists (29%, 45% and 35% respectively) would find it essential to use such tools. On the contrary, those researchers, who have given purely intrinsic motifs why they collaborate, like educating a student, fun and to learn a technology, belong to the biggest groups of project management tool users.

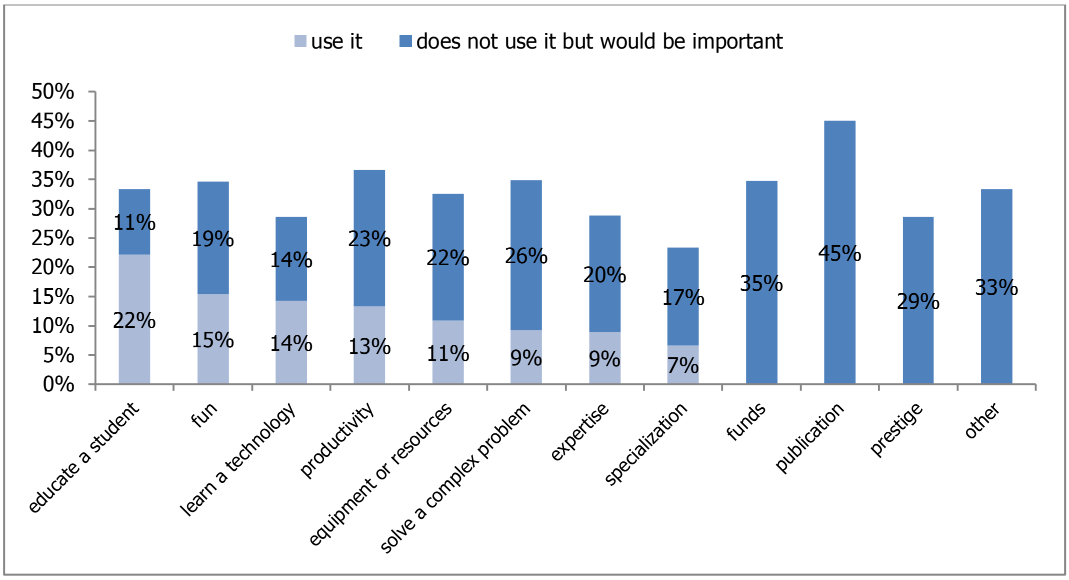

Figure 25. Scientists who use project management tools have intrinsic motivations for collaborating

7.3.5 Data management practices

Scientists found data management essential. 22% of them agreed while 78% strongly agreed that saving data is valuable for later use of for validation purposes.

Scientists used versatile and several methods to save their data. As it is visible in **Figure 26,** 83% of the asked researchers saved their data at their own PC, 72% were backing their results up regularly. 65% saved their results on the institutional server. Almost half of them (52%) used paper-based laboratory notebooks while only one out of five (20%) has moved from the paper- based notebook to an electronic system. Roughly one out of four researchers uses advanced ICT tools to save their data, like cloud-based systems (28%) or upload their data to different repositories (24%).

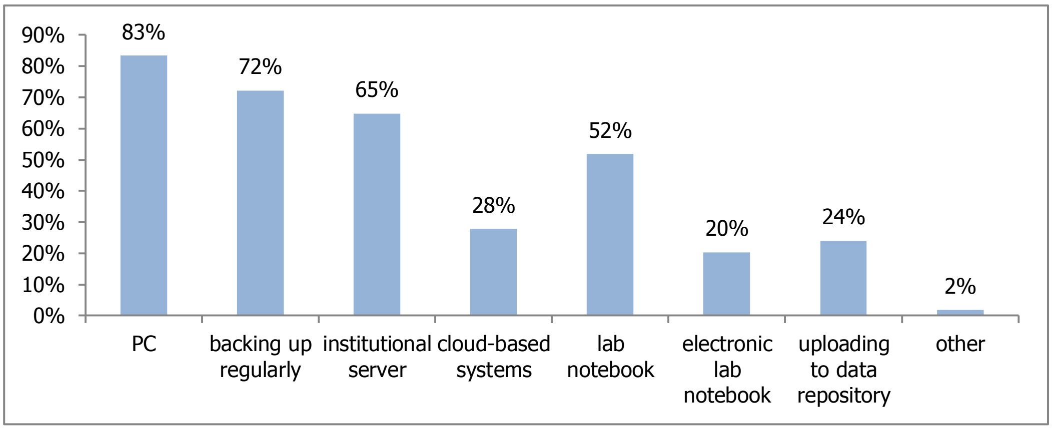

Figure 26. The ways scientists save their data are versatile

Those scientists who were participating in collaborations *to learn tacit knowledge about a technology* turned to be the group who used electronic lab books much more (35%) to save data, than the average (22%) or than, for example, those scientists who collaborated to enhance productivity (13%) (data not shown).

When it came to sharing data with the collaborating partners, scientists used mostly old-fashioned ways. **Figure 27** lists all the tools used to share data within a collaboration setup. 39% of the researchers used email, 26% used Dropbox, 6% google docs, 22% shared access to the server where the data were saved while 4% shared the electronic laboratory notebook with the collaborating partner.

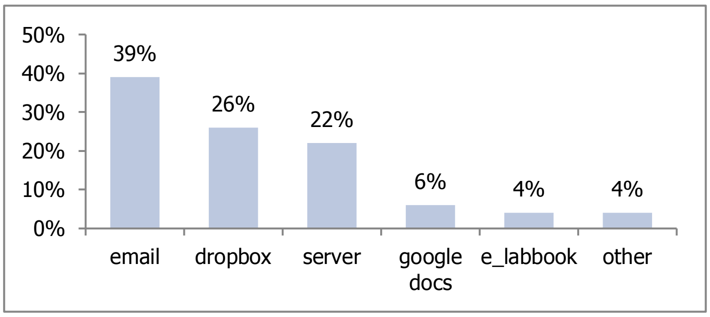

Figure 27. Data sharing practices of life scientists

Although 83% of the researchers reuse data in distinct forms, they are not used to add metadata to their data. As it is shown in **Figure 28**, **roughly half of the researchers (56%) did not use metadata** when saving their results, although some would find it valuable, 35% used metadata when saving their results, and 9% when uploading them into data repositories.

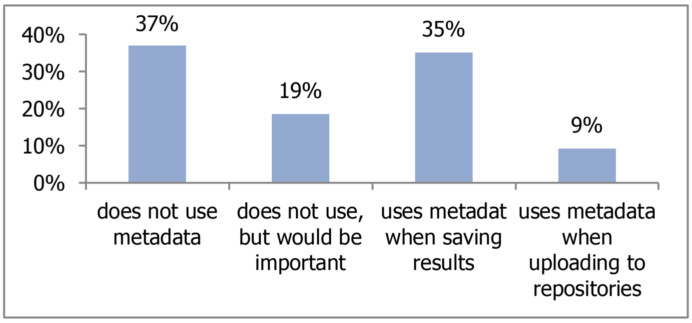

Figure 28. Less than half of the scientists use metadata

Interestingly, there was a strong significant correlation (p=0.002) between metadata and workflow management tool usage. As is can be seen in **Figure 29**, those, who were using workflow management tools (9% of the total respondents, and 26% of those who use metadata when saving results), were all using metadata when saving their data. From those who used metadata when saving data, an additional 37% thought that it would be important to use workflow systems.

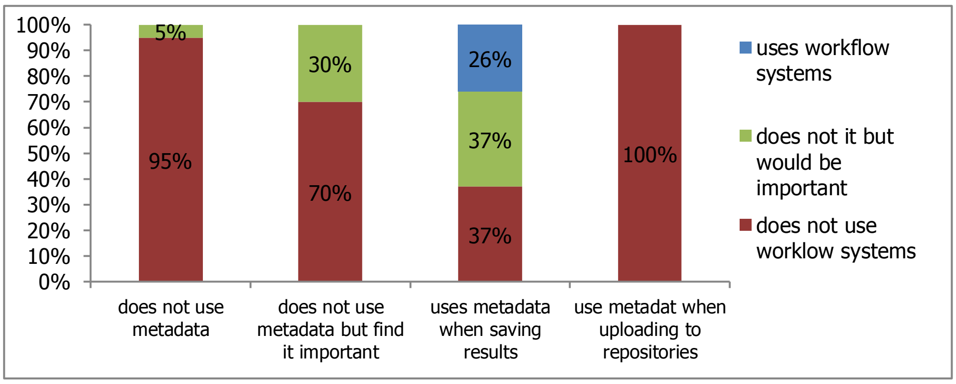

Figure 29 Scientists who use workflow systems, all use metadata when saving data

7.4 Collaboration readiness

Scientists found scientific collaborations essential for scientific research. 24% of the scientists replied that they agreed with the statement *Collaborations are essential for scientific research*, while 76% strongly agreed with it. They defined collaborations in different ways, but in principle these answers were positive. **Table 10** lists some examples of the scientists' answers.

Brainstorming, complementing each other's skills and resources.
Complementing each other's technical and theoretical bases in order to catalyse a new research idea and investigate it.
Working together with other scientists with different points of view and different expertise to solve a common problem.
Collaboration between companies and university for example for discovery of drugs
Benefiting from the strengths of both parties to make good science.
Publications, blogs, articles, emails, even chat messages, phone calls, meetings, conferences, video calls, etc. every kind of contact methods. SHARE the results, methods, work together to solve problems
Split the work, share data, combine knowledge and observations, scientific discussion
The only way to accomplish multidisciplinary research

Table 10. Scientists' definition of collaboration

The majority of life science researchers participated in collaborations. 87% of the life scientists who completed the survey (54 out of 62) participated in any kind of collaborations. (Later on, these 54 data sets were analysed further.) The distribution of the participation in intragroup, intra-institutional, national and international collaboration was almost equal (70%, 67%, 61% and 69%).

While Ph.D. students and team members mostly collaborated within their group or with other labs within the country; researchers, postdocs, senior researchers, PIs and team leaders equally participated in intra- and inter-institutional and international collaborations. The picture is clearer if we have a look at the age of scientists: based on the survey, below 25 years scientists did not take part in national or international collaborations.

67% of the respondents were participants in scientific collaborations, whereas 33% were fulfilling scientific, administrative or financial leader position in at least one of the collaborations.

7.4.1 Virtual research group

Less than half of the researchers (41%) considered their collaboration as a virtual research group. Those scientists, who thought about their collaborations as virtual research groups, had a higher tendency to agree with the statement that virtual research environments are useful for scientific research (73%), compared to those who did not see their collaborations as a virtual community (22%) ($p=0.004$), as it is visible in **Figure 30**.

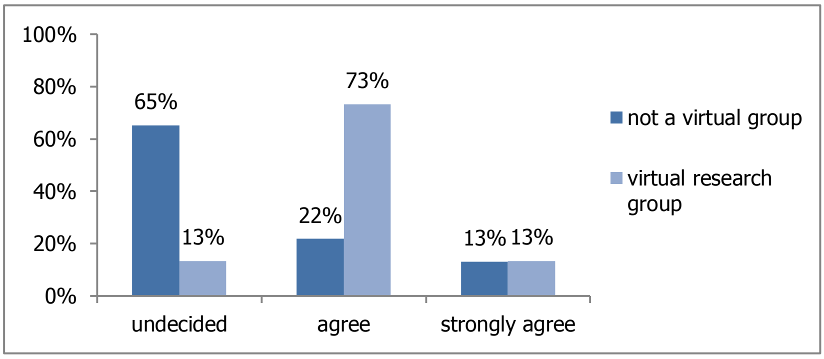

Figure 30: Researchers who consider their collaboration as a virtual research group find VREs more useful

Due to the multiple choice nature of the question, it was not possible to test the null hypothesis with Chi-square test, but as it can be seen in **Figure 31,** there was a visible increase in the number of scientists who saw their collaboration as a virtual research group within those who collaborated internationally. Roughly the half of the researchers (48%) said they saw their collaboration as a virtual research group in this group, compared to the other groups, where roughly one-third (35%, 32% and 29%, respectively) of the respondents gave this answer.

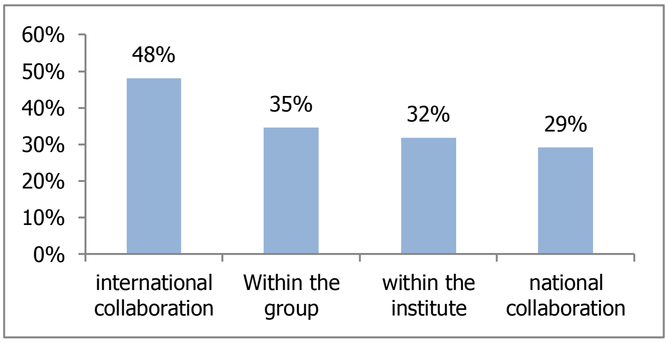

Figure 31. Scientists in international collaboration see more likely their collaborations as a virtual research group

7.4.2 Motivation for collaboration

For most of the scientists, as it can be seen in **Figure 32**, the reasons for working together were to acquire access to equipment or resources (85%), to gain access to expertise (83%), and to pool knowledge to solve a complex problem (80%). Interestingly, fun was also a relatively popular answer with 48%, just like to improve access to funds (43%), to enhance productivity (56%) and the increased specification of science (56%).

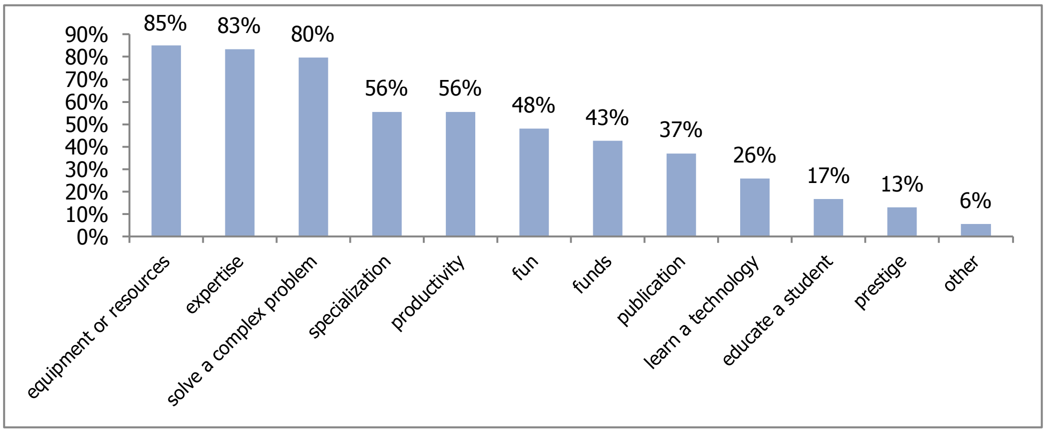

Figure 32. Scientists' motivations for collaboration

The obvious extrinsic motivation factors like prestige and publication were chosen by a relatively small number of scientists (13% and 37%), just like some intrinsic ones, like to learn tacit knowledge about technology (25%) or to educate a student (17%). Interestingly, those scientists, who named extrinsic motivation as reasons to collaborate were overrepresented in that group, which clicked important for the question *Collaborations are important for scientific research.* 57% of the scientists, who collaborate for prestige and 35% of scientists working together to gain access to publication said necessary while this ratio was only 22% for the whole group.

The motivations for collaboration showed differences in the age groups. In one hand, while collaborating to gain access to equipment and resources was one of the major reasons, it was not the case for scientists under 25 years, and funding was a reason for collaboration especially for the older age groups (45-54 and 55-64). To increase the chances of publication, on the other hand, was predominantly given as the motivation for collaboration by the younger scientists, but it showed negative tendency by the increase of age. Finally, pooling knowledge to solve complex problems was specific to the age groups from 25 to 54 (data not shown).

There was a difference in the attitude towards VREs between the different motivation groups. Those scientists who collaborate *to get access to publications*, find VREs more useful for scientist research then other scientists (they score 4 on the Likert scale, compared to the average score 3.7).

7.4.3 Collaboration type

There were **almost twice as many scientists who claimed that they were participating in integrative collaboration** (57%) compared to those who take part in sequential teamwork (37%).

There was a statistically significant (p=0.001) **correlation** between the **role of scientists in collaboration and the type of collaboration.** As it is visible in **Figure 33**, 74% of the collaboration leaders claimed that their collaboration was integrative, while only 59% of the participants belonged to the same collaboration type.

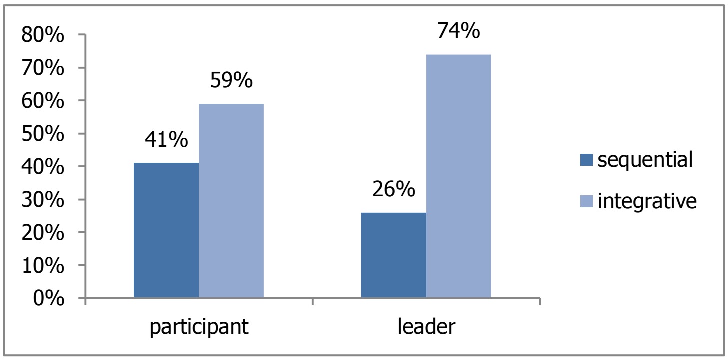

Figure 33. The majority of the collaboration leaders are in integrative collaborations

Collaboration type also had an effect on how researchers think about data management. There was a statistically significant (p=0.035) correlation between collaboration type and how relevant scientists find it to save data. As it is visible in **Figure 34**, almost all of the researchers (91%) participating in integrative collaborations strongly agreed with the statement: *It is very important to save the data for later use or for validation purposes*, while only 63% of the sequential collaborators answered the same way, and more than the third of them only agreed to it.

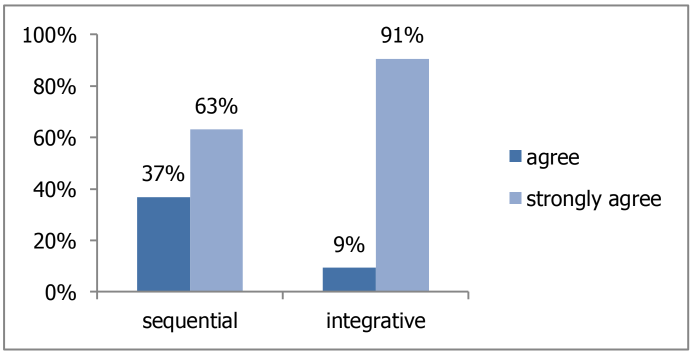

Figure 34. Scientists in integrative collaboration find it more important to save data for later use or validation purposes

7.5 Conclusions

To find out that from those obstacles, which were obtained from the analysis of the interviews made with life science researchers, which ones are relevant for a larger group, I have performed an online interview. Most of the scientists have never used a VRE, and the majority have not even heard about this infrastructure. From the 54 life science researchers filling out the survey completely, only 3 have heard about virtual research environments, while 7 were familiar with one of the other terms. From these 10 researchers, only 6 used this kind of infrastructure before. Those who shared their experiences, all had difficulties with using VREs.

Scientists had a highly diverse attitude towards an EU-wide VRE, which can be interpreted as **likeability** of the VREs. The major concerns about a huge infrastructure were related to the increase in bureaucracy, the size of the infrastructure (too big) and its efficiency, data standards, data safety and security, legal and ethical issues related to patient data, equal access, and finally sharing data with competitors. Some scientists were mentioning that some other things would be more relevant, like building a data repository or forcing open access publication of research funded by public money. One participant was against the concept because it would not be voluntary. These concerns could be taken into account during the marketing of such large-scale VREs. From the positive side, the following arguments were mentioned: a place to publish negative results and detailed protocols, data would be freely available to the public, beneficial for collaborations. Scientists had various thoughts about the **usefulness** of a VRE. 41% of the scientists were not sure whether VREs could be beneficial for scientific research while in total 57% found that virtual research environments could be useful for scientific purposes (agreed and strongly agreed with the related statement).

The respondents communicated by using multiple **communication means**. They have used email and had meetings face-to-face. Skype was less frequently used, only in the 65% of the cases, and only 1 out of 10 scientists used social media to communicate with their collaborating fellow scientists. More than half of the interviewees were satisfied with their current communication practices. One fifth of them mentioned that they would like to have more frequently face-to-face meetings. Others suggested qualitative changes, such as reducing the number of tools, to have one platform for oral and written communication, to have a system which stores all communication related to a single project in an organized way, or to embeds attachments or metadata into emails.

Roughly half of the scientists in the survey, just like half of the interviewees, were participating in **science outreach activities**. The survey supported the result of the interviews performed with life science researchers, namely that science outreach is dependent on the mindset of research group leader. Scientists bearing any kind of leader position in collaborations or in research groups were more likely to participate in science outreach activities. Interestingly, there was no correlation between the use of social media (general or academic) and the engagement in science outreach activities, as it would have been expected from the latest PEW thesis (2015). This thesis concluded that 47% of their survey participants were using social media to talk about science or read about scientific discoveries and 24% of them wrote blogs about scientific issues (PEW 2015). This suggests that scientists, who were active in science outreach in the VRE survey, did not do it (exclusively) via social media, and scientists who used social media, did not use it (exclusively) for science outreach. This is represented as well in the list of activities scientists mentioned as science outreach. Besides writing blogs, popular science papers and managing a social media website open to the general public, a lot of offline activities were mentioned, such as organizing science-related events for students or for the general public, workshops and science week, in which SNSs were probably not involved.

While almost all scientists found networking necessary, they had very diverse opinions about **social networking sites** and their role in the academic world. In total, 18 scientists wrote down that social media was or could be useful, with or without additional concerns. The most common concern about social media was that it is only good for personal, but not for professional purposes. A few people mentioned additional concerns about privacy, copyright and data safety, or that social media can be dangerous or especially harmful for science, that the facts are not checked, and that it consumes too much time. This is in line with previous studies, which showed that many scientists see Facebook and Twitter unsuitable for professional use, just consumes too much time (Van Eperen and Marincola 2011), or even dangerous (Procter, Williams et al. 2010). Those scientists who mentioned positive aspects of social media were writing that it is useful to public engagement, education or science outreach, for networking, for funding, for collaboration, for finding and disseminating information, to follow research progress or to build a public profile. One scientist mentioned that not scientists, but particular people should deal with social media. These results reassure why there was no correlation between science outreach and social media use – only 5 scientists mentioned that they used social media for this purpose.

Interestingly social media can generate intense feelings in users, such as hate or love, or can annoy people. The attitude towards social media showed age-distribution. The youngest age group was rather very positive or had mixed feelings towards social media. By the increase of age, more negative thoughts about social media sites appeared. The versatile attitude was represented in the actual social media use as well. 11% of the researchers answered that they did not use social media at all. Almost the half of the researchers (47%) used social media only for personal purposes while 43% for academic purposes. This percentage is comparable with results of the previously mentioned study, in which 45% of the asked scientists used web2.0 for scientific purposes (Van Eperen and Marincola 2011). On the margin, those scientists, who used social media for academic purposes, were more positive about social media than those, who only used these sites for personal purposes or did not use it.

Comparable with the findings of Nature's online survey (Van Noorden 2014), in which roughly 90% of the participants used **science-related social networks**, like Research Gate, Mendeley, ORCID, Academia.edu and so on, 88% of the respondents to the VRE survey used academia-related SNSs. The 88% is much higher than the percentage of those scientists who answered to the previous question that they were using social media for academic purposes (43%). This discrepancy is arising from the fact, that half of the scientists who claimed that they did not use social media at all were actually using social networking sites designed for a scientist, just like 88% of those, who said that they were using social media only for personal purposes. This suggests that scientists do not consider SNSs designed for academic use as social media. Only 17% of the academic SNS-users were active users; the rest were just using these sites passively. There was a significant correlation between the position of scientists in collaboration and the active use of academic social networks. The majority of those who do not use any kind of academic-related social networks had only participatory role in collaborations, while 26% of collaboration leaders were active users of these sites, compared to the 12% in the case of collaboration participants.

The results of the survey suggest a relationship between social media use and the attitude towards VREs. While in the group of scientists, who claim that they do not use social media at all, there was no one with a slightly positive or absolutely positive attitude towards VREs, those who used social media for private or for academic purposes had a more diverse opinion regarding VREs, ranging from the sceptical to the positive. Interestingly, 44% of those scientists, who use social media only for their personal purposes, had an absolutely positive attitude towards VREs, while this was only 21% in the case of scientists using social media for academic purposes as well.

Roughly one out of ten researchers used any kinds of **project management tools**, and roughly one out of five thought that it would be important to use them. Although the two tools provide solutions to entirely different problems, the distribution of scientists using project management and workflow management tools looked almost the same. The users of these tools were not completely overlapping, so not only those scientists used project management tools, who used workflow management systems, so this gives an indication of the general ICT-tool use of scientists. There was a difference between how collaboration leaders and participants thought about project management tools. Almost twice as much collaboration leader used project management tools then collaboration participants, and half of those leaders, who did not use these tools, found it rather important to use them. This is not surprising if we take into account that scientific, administrative or financial leaders in collaborations face much more often organisational issues then those scientists, who only participate in collaborations. What is more interesting is that those scientists, who mentioned enhancing productivity as one of the motivations for collaborating used project management tools the most. On the contrary, none of the scientists who collaborated for prestige used project management tools, although half of them found it important.

All of the researchers found **data saving** for later use of for validation purposes relevant. Scientists used various methods to save their data. The majority of them saved their data at their own PC, and most of them were backing their results up regularly. Roughly two-thirds saved their results on institutional servers. Almost half of them used paper-based laboratory notebooks while only one out of five has moved to an electronic notebook system. Almost the same amount (20%) of scientists used other advanced ICT tools, like cloud-based systems and uploaded data to repositories. Just for comparison, 10% scientists used workflow or project management tools. Those scientists who were participating in collaboration to learn tacit knowledge about a technology turned to be the group who used electronic lab books much more to save data than the average or the other groups. This suggests that those scientists who were more open to learning new scientific technologies were more open to new science-related ICT tools too.

Although the vast majority of researchers reused data in distinct forms, roughly half of them did not use **metadata**. One-third of the scientists used metadata when saving their results, and only one out of then when uploading them into data repositories. Interestingly, there was a correlation between metadata and workflow management tool usage. Those, who were using workflow management tools, were all using metadata when saving their data. When it came to sharing data with the collaborating partners, scientists used mostly old-fashioned way. The most frequently used tool was email, which represents the 1.0 way of managing and sharing information (Bartling and Friesike 2014). Dropbox, Google docs, and shared access to the server represent the 2.0 or cloud way by providing collaborative writing (Bartling and Friesike 2014). These 2.0 tools were not that predominantly used.

Every scientist found **collaboration** essential to scientific research, and the majority (87%) of life science researchers who filled out the survey, participated in collaborations. The distribution of the participation in intragroup, intra-institutional, national and international collaboration is almost equal. While Ph.D. students and team members mostly collaborated within the group or with other labs within the country; researchers, postdocs, senior researchers, PIs and team leaders equally participate in intra- and inter-institutional and international collaborations. The picture is clearer if we have a look at the age of scientists: below 25 years scientists did not take part in national or international collaborations. Ph.D. students do not have time, skills and the expertise to participate in huge projects unless their project is specially defined as being a part of a consortium project by their supervisors. Collaborations are also one indication of achievement and acceptance. As scientists become more experienced, they become more independent as well to establish their own collaborations (Bozeman and Corley 2004). One-third of the respondent was fulfilling scientific, administrative or financial leader position in at least one of the collaborations. Not surprisingly there was a correlation between the function of a scientist in the group and their role in collaborations. PIs, team leaders were more frequently leading collaboration then Ph.D. students or postdocs.

Less than half of the researchers considered their collaboration as a virtual research group. Those scientists, who did, had a higher tendency to find virtual research environments useful for scientific research. The largest percentage of scientists thinking about their collaborations as virtual research group was in the group who participated in international collaborations.

For most of the scientists, the most frequent reasons for working together were to acquire access to equipment or resources, access to expertise, and to pool knowledge to solve a complex problem. Interestingly, fun was also a relatively popular answer, just like to improve access to funds, enhancing productivity and increased specification of science. The obviously extrinsic motivation factors like prestige and publication were chosen by only a few scientists. The motivations for collaboration showed differences in the age groups. On the one hand, while collaborating to gain access to equipment and resources was one of the major reasons for most age groups, it was not the case for scientists under 25 years, and funding was a reason for collaboration especially for the older age groups, above 45 years. To increase the chances of publication, on the other hand, was predominantly given as the motivation for working together by the younger scientists, and it showed a decreasing tendency with the increase of age. Those scientists who collaborated to get access to publications found VREs more useful for scientific research than the other scientists.

There were almost twice as many scientists who claimed that they were participating in integrative collaboration compared to those who took part in sequential collaboration. There was a correlation between the role of scientists in collaboration and the **type of collaboration**. Three-quarters of the collaboration leaders claimed that their collaboration was integrative while only 59% of the participants belonged to the same collaboration type. Collaboration type also had an effect on how researchers thought about data management: researchers participating in integrative collaborations found more important to save data for later use or for validation purposes.

8 Conclusions

The aim of this project was to find out what obstacles prevent life science researchers from adopting virtual research environments, by doing a literature review, interviewing VRE researchers and developers and potential users, and doing an online survey among potential VRE users, and finally to build a virtual collaboration readiness model from these result.

To investigate the obstacles, I have used a mixed method. First I have performed a detailed literature review to explore the factors previous studies have determined. Based on these publications and related theories I have created a theoretical framework, which guided me in the further parts of the research. It helped me to perform the interviews with VRE developers and potential users, as well as in the creation of the online survey to involve a larger number of life scientists into the research.

In this Chapter, I summarize the results of the research by listing the key factors under the dimensions of collaboration infrastructure readiness, collaboration technology readiness and collaboration readiness (Olson, Teasley et al. 2002) by using the theoretical framework (from Chapter 4.2) as a structural base.

8.1 Collaboration infrastructure readiness

Virtual research environments are ICT tools that are designed to assist the various aspects of scientific collaborations. The infrastructure has to be adequate and function properly in the given virtual environment to have a working collaboration tool (Olson, Teasley et al. 2002). But above the technical details of the technology itself, there are some other aspects of the infrastructure which are crucial in influencing the attitude of scientists towards VREs.

The **definition of VREs** for VRE researchers and developers was rather plastic; almost all of the interviewed experts had a different interpretation. Half of the interviewed researchers were **not aware of what VREs were**, even if they were active users of the we-NMR, which is a VRE. Most of the scientists filling out the survey (90%) have never used a VRE, and the majority (80%) have not heard about them, even not under different terms. Those who shared their experiences had all difficulties with using VREs.

Shackel (1991) disseminated three main attributes of the infrastructure that must balance the costs of the tool (including financial costs, as well as social and organisational consequences) to lead to acceptance: likeability, utility and usability (Shackel 1991). Regarding **likeability**, all of the interviewed researchers were positive about the idea of building a huge EU-scale VRE to serve as a data repository and an SNS for scientists. They indicated that such a VRE could help keeping the data accessible, make science more democratic, transparent and increase efficiency, which is parallel to previous research results (Dutton and Jeffreys 2010). Those scientists, who were not aware of what VREs were, had a rather positive attitude this kind of infrastructure and mentioned only learnability as a potential problem for PIs. On the contrary, scientists, who had hands-on experience with VREs were aware of their potential and actual problems.

The second attribute, **utility** refers to the match between user needs and product functionality (Shackel 1991). It is a general problem that the actual service implemented in

VREs is not what the users needed (Carusi and Reimer 2010). VRE researchers and developers had a diverse opinion about the utility of VREs. Some claimed that scientists do not need VREs while others were convinced of the opposite, but all R&D professionals agreed that there was a need for integrative, interoperable infrastructure, which should help scientists in saving and sharing data. One developer has mentioned a potential solution to the problem of utility: **cooperative co-design**, which is line with the user-centred design suggested by Olson and his colleagues as well (Olson, Teasley et al. 2002).

The third attribute of infrastructure is **usability** which refers to users' ability to utilise the functionality in practice. Usability tests are measures how successfully users can reach some particular goals by using the tool. Ease of use (including flexibility and learnability), effectivity and finally satisfaction (success rate in reaching a goal) determine the actual user's perception of the usability of the product (Shackel 1991). The key factors (in terms of usability) in adopting VREs mentioned by VRE developers and potential users were: standardization, interoperability, integration, compatibility, single-sign-on system, learnability and flexibility, which factors came across during the literature search (Carusi and Reimer 2010, van der Vaart 2010).

8.2 Collaboration technology readiness

The basic functions of a standard virtual research environment contain ICT tools that provide a secure and safe environment, with federated authentication, personalized access control, user management and user statistics. These functions serve as the basis for the VRE, other function, are built on these (for an overview, please check **Figure 5** on page 27). An average VRE also contains **document organization and sharing tools**, archives, search facilities, **communication tools**, which provide possibilities to communicate synchronously and asynchronously and to share information; **project management tools** which provide help in distribution and completion of collaboration- or research-related tasks. Research registration, alerts, manuscript submission and review systems and reference management are useful **extra applications** for scientists, as well as the functionality of migrating data to external repositories after publication. Other extra functionalities as shared research applications, electronic lab notebooks, and shared instrumentation can increase the utility of an environment. Finally VREs, especially the bigger ones, can have specialized subsites, for example for different communities, public sites for science outreach or conference sites (Allan 2009).

8.2.1 Communication

Communication is a fundamental part of all stages of collaborations. Without proper communications, tasks could not be synchronised; scientists would not learn from each other, results would not be integrated, and it would be difficult to keep trust (Sonnenwald, Solomon et al. 2002). Communication also plays a crucial role in the development of the process of forming of virtual communities (Rheingold 1993). Many of the modern ICT tools can enhance collaborations, but these tools by themselves cannot guarantee effective working together. The people who are actually doing the collaboration are still crucial to maintain a healthy relationship, do the distributed tasks and communicate with each other. With good and client-specific design, ICT tools can counteract some of those human tendencies that can weaken collaborations. Research has been shown that an online platform for collaboration providing only support for communication between the members worked nicely only until any conflict occurred in the group. After a confronting situation, the efficiency of this platform was decreasing due to more and more negative interactions and less and less member satisfaction. These kinds of situations could be better solved with the introduction of tools that can help the process of conflict management, like a multi-criteria decision analysis tool or a problem formulation routine. In light of this information, it is interesting to see that most virtual research environments help only communication without providing support for problem solving (Poole 2009).

Life scientists communicated by using multiple **communication means,** mostly by face-to-face or by email and Skype. Only 1 out of 10 scientists used social media to contact their collaborating fellow scientists. One of the interviewed scientists was quite sceptical about computer-mediated communication for professional use; she preferred personal communications. Some previous research explains the requires for the high rate of face-to-face interaction among the scientists with the informal nature of collaboration formation (Bozeman and Corley 2004), or that ICT tools can cause obstacles in the information exchange in collaborations (Qureshi, Min et al. 2005). Sonnewald and his colleagues (2002) found that the introduction of video conferencing and an electronic whiteboard during group meetings increased the level of formality, and reduced the effectiveness of the meetings.

They could enhance the efficiency of the discussions by adding and supporting informal interactions (Sonnenwald, Solomon et al. 2002).

All of the interviewed, and more than half of the surveyed scientists were satisfied with the communications tools they used. Some suggested quantitative changes in communication within collaboration (more meetings), others suggested qualitative changes, such as reducing the number of tools, to use one platform for oral and written communication, to have a system which stores all communication related to a single project in an organized way, or to embeds attachments or metadata into emails. One researcher mentioned that she did not like the current communication ways she had with her colleagues because these meetings were too official and not fun anymore. Not only were the interviewees, in general, satisfied with the used communication tools, but most of them mentioned that they also needed these tools to maintain a connection with the collaborating partners.

8.2.2 Science outreach

With the wide-spread use of social media, blogs and wikis, scientists can explain their scientist results, and the societal - ethical implication of these findings directly to a wider audience without the filter of the conventional media. Due the interactivity of the web2.0 and social media sites, there is a possibility to start a discussion, so these ICT tools are found to be ideal **for public engagement**. The traditional **scholarly communication** is also changing thanks to social media. The majority of scientific papers, conference summaries, and poster abstracts are published mostly digitally and spread or publicized by science blogs and social media sites (Bartling and Friesike 2014).

Roughly half of the scientists in the survey, just like half of the interviewees, were participating in **science outreach activities**. The survey supported the result of the interviews performed with life science researchers, that science outreach is dependent on the mindset of research group leaders. Scientists with a leader position in collaborations or in research groups were more likely to participate in science outreach activities. Interestingly, there was no correlation between the use of social media (general or academic) and the engagement in science outreach activities, as it would have been expected from the latest PEW thesis (2015). This suggests that scientists, who were active in science outreach, did not do it (exclusively) via use social media. Besides writing blogs, popular science papers and managing a social media website open to the general public, a lot of offline activities were mentioned as science outreach activities, such as organizing science-related events for students or for the general public, workshops and science weeks, in which SNSs were probably not involved.

8.2.3 Social networking sites

Virtual research environments have a function, which can be categorized as a social networking site (SNS). These sites require that the members create a profile, ten add connections to other members, creating a list of associations, and finally, users can navigate through such associations to access a wider network (Orchard, Fullwood et al. 2014). Besides the general sites, there are more and more specialized social networking sites also for academics, like LabSpaces, Academia, Mendeley, ResearchGate, Surgytec, Sciatble and so on (Van Eperen and Marincola 2011). These academic networking sites were shown in various studies to facilitate communication between scholars and enhance scientific collaboration (Jordan 2014).

While almost all scientists found networking necessary, they had very diverse **opinions** about social networking sites and their role in the academic communication. In this study, scientists rarely mentioned that they liked or disliked any kind of computer-mediated communication but social media generated intense feelings in users, such as hate, love or annoyance. The most common concern about social media was that it is only good for personal, but not for professional purposes. This is in line with previous studies, which showed that many scientists see social networking sites unprofessional, and in general that it just consumes too much time (Van Eperen and Marincola 2011). The attitude towards social media showed age-distribution. The youngest age group was rather very positive or had mixed feelings towards social media. By the increase of age, more negative thoughts about social media sites appeared. Those scientists who mentioned positive aspects of social media were writing that it is useful to public engagement, education or science outreach, for networking, for funding and collaboration, for finding and disseminating information, to follow research progress or to build a public profile.

The versatile attitude was represented in the actual social media use as well. One out of then researchers answered that they did not use social media at all. Almost the half of them who used social media, used it only for personal purposes and a bit less for academic purposes, which is comparable to the results of previous studies (Van Eperen and Marincola 2011). Those scientists, who used social media for academic purposes, were more positive about social media than those, who only used these sites for personal purposes or did not use it. Comparable with the findings of Nature's online survey (Van Noorden 2014), 88% of the respondents to the VRE survey used academia-related SNSs. This percentage is almost double of the percentage of those scientists who answered to the previous question that they were using social media for academic purposes. This is due to the fact, that half of the scientists who claimed that they did not use social media at all were actually admitting using social networking sites designed for a scientist, just like 88% of those, who said that they were using social only for personal purposes. This suggests that scientists do not consider SNSs designed for academic use as social media. Only 17% of the academic SNS-users scientists were active users; the rest were just using these sites passively.

Procter and colleagues stated that there was a positive association between the degree of adoption of web2.0 tools and the involvement of scientists in collaborative research. Those who were working with different institutions (interinstitutional collaboration) were more likely to be frequent users of web 2.0 tools (Procter, Williams et al. 2010). In this particular research, there was no correlation found between the scope collaboration and the use of social networking sites.

The results of the survey suggested a correlation between social media use and the attitude towards VREs. While in the group of scientists, who claim that they do not use social media at all, there is no one with a slightly positive or absolutely positive attitude towards VREs, those who use social media for private or for academic purposes have a more diverse opinion regarding VREs, ranging from the sceptical to the positive

8.2.4 Project management practices

Collaborations are very complex projects regarding project management issues. First, some of the collaborating partners are from another country, which can cause several problems in terms of language use and cultural differences. Second, scientific research is a risky business, the outcome of a research project cannot be promised (Beukers 2011).

The opinions of different VRE developers were contradictory on whose role is it to deal with project management. One developer was concerned that scientists needed project management tools while another saw project management as a service, which should be provided to scientists. A third VRE expert thought that the independent scientific mindset is an obstacle in using standardized project management tools in academia.

From the researchers' side, two out of the eight interviewees used an ICT-based project management tool in their lab. One of them agreed that his group members had difficulties in adopting this tool, and the reason for this is that it is hard to change someone's habits, reassuring that VRE developer's opinion, that scientists have to change their mindset to use these kinds of tools. Roughly one out of ten researchers from the VRE survey used any kinds of project management tools, and roughly one out of five thought that it would be important to use them. This is in line with the previous findings, which state that although in life sciences, international and interdisciplinary collaborations all have to struggle with difficult project management tasks, the importance of project management is not acknowledged (Allan 2009, van der Vaart 2010).

8.2.5 Data management practices

In the past few decades, life sciences became more and more data-centred. Scientific experiments are generating huge amounts of data, and also cost a lot of money. Therefore it was suggested to share as much data as possible for reuse by further research. The reuse of data requires not only extensive and proper data management practices but new attitude from life science researchers as well.

All VRE R&D professionals agreed that data management systems were crucial for scientific research and for collaborations. One researcher even added that the adoption of VREs is dependent on the adoption of data management systems. Another researcher mentioned that VREs or Science Gateways could be built on data repositories because these remote data storage places are used by data analysis software or tools, which can be part of a science portal or a virtual research environment. These ideas are similar to what which Kaizer and Kuiper suggested, that advanced data management is required to see the added values of VREs (Kaizer and Kuiper 2014). Regarding data management and uploading data after publication into open repositories, institutional and governmental/EU-research policies are crucial in giving directions but also in motivating researchers for example with an extended citation index system. This view is in line with what the measurement school of the e-research stands for (Bartling and Friesike 2014).

From the scientists' side, all of the researchers found data saving for later use of for validation purposes relevant. Scientists used versatile methods to save their data. The majority of them saved their data at their own PC, and most of them were backing their results up regularly. Roughly two-thirds saved their results also on the institutional server. Almost half of them used paper-based laboratory notebooks while only one out of five has moved from the paper- based notebook to an electronic system. Almost the same amount of scientists used other advanced ICT tools, like cloud-based systems and uploaded data to different repositories.

In scientific collaborations sharing data, information and knowledge the use of metadata is required. Metadata is usually accessed as a catalogue of information; it contains not only a description, how the data were generated, but also provenance (history of ownership) and location referring to the data set (Allan 2009). These kinds of information are essential for those who would like to use the data uploaded into special data storage places, data repositories. Twelve percent of the interviewed scientists mentioned that they have ever labelled their data or used metadata. This percent was around 33% in the survey respondents. When it came to sharing data with the collaborating partners, scientists used mostly old-fashioned ways. The most frequently used tool was email, which represents the 1.0 way of information sharing (Bartling and Friesike 2014). Far fewer scientists used 2.0 tools, like Dropbox, Google docs, and shared access to the server where the data were saved. Extra functions, like adding comments to files or datasets, or sharing tools were seldom used by researchers, but the majority of them found it rather important.

The results of this study support the findings of the study performed by Kaizer and Kuiper: researchers plan and perform experiments without data reuse in mind, and data management and sharing practices are usually not at that level, which could serve the reuse of experimental outcomes (Kaizer and Kuiper 2014).

8.3 Collaboration readiness

Nowadays research funding agencies, like the European Science Foundation prefer to give support to research grants of interdisciplinary, international and inter-institutional collaborations (Sonnenwald 2007). The general reasoning behind that phenomenon is that individual scientists or single research groups do not possess enough knowledge, expertise and time to perform research which is appreciated by the funding agencies and academic publishers (Hara, Solomon et al. 2003).

Every scientist participating in this research found collaboration essential to scientific research, and the majority (87%) of life science researchers who filled out the survey, participated in collaborations. The distribution of the participation in collaborations within the research group, within the institution, and international collaborations was almost equal. One-third of the VRE-survey respondent was fulfilling scientific, administrative or financial leader position in at least one of the collaborations.

8.3.1 Community aspects

The adaptation of technology, in our example, the adaptation of collaboration-assisting virtual environments, has a social component, which requires that all members of the given collaboration conforms to a common pattern of interaction and settled rules (Qureshi, Min et al. 2005). Moreover, it was shown that communities that already existed before the introduction of the cyberinfrastructure worked better in the virtual space than those, which were formed at the time of the creation of the VRE (van der Vaart 2010). Community uptake, elaboration and maintenance of social networks are central to the sustainability of the VREs as well (Olson, Teasley et al. 2002, Carusi and Reimer 2010). Therefore, community aspects have a significant role in determining scientists' attitude towards these environments.

More than half of the VRE developers mentioned that VREs or research infrastructures have **various communities**, developers, users, but also trainers and educators, and these communities have different needs regarding the use of the infrastructure. Several R&D developers agreed on, that besides scientists experts should also be included in the VREs or research infrastructures, who can give advice on particular questions. Based on the opinion of one VRE R&D professionals, the main challenge in the communication within collaboration is not the information exchange, but the engagement of the community members.

Less than half of the researchers considered their collaboration as a virtual research group. Those scientists, who considered their collaborations as **virtual research groups** had a higher tendency to find virtual research environments useful for scientific research, compared to those who do not see their collaborations as a virtual community. The largest percentage of scientists thinking about their collaborations as virtual research group was in the group who participated in international collaborations.

8.3.2 Funding

The perspective of **continued funding** is crucial in the uptake of VREs. This can be problematic with the currently running EU-projects as well. One solution to this problem, based on the interviews performed with VRE developers, is that the financing of these VREs become part of the standard institutional budget. This is similar to the idea suggested by other researchers (van der Vaart 2010). The institutional budget does not *per se* means that the universities pay the costs of the maintenance, national research institutions can also provide budget, VREs can go commercial, so members pay for the service, or SMEs, companies or the public invest in the budget of VREs. Another possibility is to use standardized protocols in VREs; everything can be transported into another infrastructure without trouble.

8.3.3 Institutional policies

Research institutions and policies related to scientific research are important in the process of adopting VREs by life science professionals. It has been shown that a supportive local environment is crucial in the acceptance and spread of web2.0 tools among scientists (Procter, Williams et al. 2010). Functions that are related to long-term infrastructure building do not fit into a typical academic position, so institutions should provide positions for people who are responsible for the development, management and governance of VREs, or organize communities for these infrastructures (Lawrence and Wilkins-Diehr 2012).

Data management recommendations are also an essential part of the institutional and research policies. These can shape the current practices of life science researchers. Based on the interviews research institutions are currently busy with making these kinds of recommendations, and PIs or group leaders are generally involved in these decisions. On the contrary, none of the postdocs and Ph.D. students I have done interviews with was aware of the institutional policies related to data management.

8.3.4 Differences in language

Interdisciplinary and international collaborations are beneficial to scientific research because these deliver broader perspective, collective expertise and higher chances of gaining grants. But the internationality and the interdisciplinary nature can also cause misunderstanding and loss of trust due to the differences in the languages and the work practices (Carusi and Reimer 2010). One VRE developer drew the attention to another aspect of using different languages – namely using contradictory definitions in various scientific disciplines. Above the non-overlapping jargons, different disciplines have different work practices as well, like using different standards for using metadata, which is not interchangeable.

There are cases, in which legal, ethical and cultural issues are the obstacles in using a VRE due to different users. A typical example is when fear of losing ownership of data and ideas stands in the way of sharing (Carusi and Reimer 2010). This was one of the major concerns of the VRE researchers that the information they share could be reached by their competitors as well and they will experience the disadvantages of sharing data. This situation has to be handled carefully, because as it has been shown, forcing these researchers to share all the data that they kept private before publishing could result in an adverse attitude towards the infrastructure itself (van der Vaart 2010).

8.3.5 Motivation for collaboration

Hara and Solomon (2003) found one of the critical factors that are important in forming a scientific collaboration is the reason why scientists want to work together. They have categorised these incentives into external (funding, publications, and prestige) and internal (learning new technologies, solving complex problems or fun) motifs. Their study showed that the external motivations were significant in sequential collaborations while integrative collaborations required internal motivations (Hara, Solomon et al. 2003).

Most of the interviewed researchers claimed that the specialization of science is the main force that drives collaboration, meaning specialization of expertise and specialization of equipment. Others mentioned that without collaboration, they could not publish their results, as scientific publishers prefer authors from different institutes. For most of the scientists who filled out the survey, the most frequent reasons for collaboration were to acquire access to equipment or resources, access to expertise, and to pool knowledge to solve a complex gain problem. Interestingly, fun was also a relatively popular answer, just like to improve access to funds, enhancing productivity and increased specification of science. The obviously extrinsic motivation factors like prestige and publication were chosen by only a few scientists. The motivations for collaboration showed differences in the age groups. In one hand, while collaborating to gain access to equipment and resources was not necessary for scientists under 25 years and funding was a reason for collaboration especially for the older age groups, above 45 years. To increase the chances of publication, on the other hand, was predominantly given as the motivation for collaboration by the younger scientists.

8.3.6 Type of collaboration

Based on interviews performed in a science centre, Hara (2003) suggested that there are two main categories of scientific collaborations: sequential and interdependent collaborations. The most general form of scientific teamwork is sequential interdependence, and as it only requires awareness and complementarity, it is rather easy to establish. Scientists in this kind of partnership work on the same project, but they are responsible for their parts of the project. In the end, they provide the results of these sub-projects, and the result is finally bigger than any member could accomplish on its own. This kind of collaboration is frequent in the field of chemistry and biology (Walsh and Bayma 1996), due to the complementary expertise of the scientists. Collaborations in which participants work together throughout the whole research process, from developing ideas, through carrying out the research and solving occurring challenges, to summing up the results are called as integrative collaborations. This type of collaboration requires more respect and trust, and can lead to more conflicts over responsibility and contribution (Hara, Solomon et al. 2003). In the survey, there were almost twice as many scientists who claimed that they were involved in integrative collaboration compared to those who take part in sequential collaboration. Three-quarters of the collaboration leaders claimed that their collaboration was integrative while only 59% of the participants belonged to the same collaboration type. Collaboration type also had an effect on how researchers think about data management: researchers participating in integrative collaborations found more important to save data for later use or for validation purposes.

Procter and colleagues stated that there was a positive association between the degree of adoption of web2.0 tools and the involvement of scientists in collaborative research. Those who were working with different institutions were more likely to be frequent users of web 2.0 tools (Procter, Williams et al. 2010). This is line with the result of the VRE survey, which showed that scientists who considered their collaborations as virtual research groups had a higher tendency to find virtual research environments useful for scientific research, and that the maximum percentage of scientists thinking about their collaborations as virtual research team was in the group who participated in international collaborations.

One interviewed PI had experienced that the sequential collaborations were more successful than those, in which scientists were more interdependent, these did not accomplish what they have planned. Interestingly, most of the structural biologists were involved in sequential collaborations while the biomedical scientists were developing the consortium together, showing a discipline-specific difference.

8.3.7 Disciplines

Scientists from different scientific fields collaborate differently. There are disciplines, which are famous for their high tendency to work together. For example, physicists collaborate more than social scientists. Besides the historical aspects and the external factors that force scientists to work together (like the price and availability of the research tool or infrastructure), there are other factors which are discipline-specific, and can affect the motivations of scientists to collaborate. The data they deal with during the research is also different, so the way they share and analyse these data is also different. Scientists from various disciplines have also found to have different priorities regarding the importance of the various elements of the research lifecycle. For a huge percentage of researchers doing biological science, it is not important to do deal with the managements of research outputs (Allan 2009). Based on these differences it was suggested, that different disciplines would adopt VREs with varying rates (Sonnenwald, Solomon et al. 2002, Allan 2009).

Experts highlighted in many cases that there are differences in the uptake of VREs between disciplines. The major distinction was made between the great scientific domains, like humanities and social sciences versus life sciences, but some experts mentioned differences between life science disciplines as well, which is in line with what the literature suggests (Sonnenwald, Solomon et al. 2002, Allan 2009, Markauskaite, Kennan et al. 2012). The interview results showed the type of collaboration can be dependent on disciplines.

9 Building a model for virtual collaboration readiness

To answer the fourth sub-question, which is focusing on how it possible to overcome the obstacles that were found to stop life science researchers in adopting and using VREs, I have built a model for virtual collaboration readiness, which is visualized in **Figure 35**. The model is based on the theoretical framework, which is the combination of the literature review and topic-related theories, such as the technology acceptance model and the network theory. The interviews performed with VRE R&D professionals and scientists, plus the survey results added information in defining the current situation and the further steps to overcome the obstacles.

The model is divided into three parts, which represent the three dimensions of the theoretical framework. The core of this model is the technology acceptance model. The three dimensions, as described in more details in Chapter 4.2, represent the three attributes of the TAM. Collaboration readiness can be understood as the collection of external factors that influence the perceived usefulness and perceived ease of use. In this dimension, I have listed in community-related factors, which were shown to be influential in a virtual collaboration setting. Collaboration infrastructure readiness, which focuses on the technological aspects of the adoption of VREs, corresponds to the perceived ease of use in the TAM; as the technical details and other infrastructure-related factors crucially determine how users perceive the given infrastructure. Finally, collaboration technology readiness represents the perceived usefulness of VREs. This dimension is focusing on the individual scientists' practices in terms of communication, project management and data management.

In this chapter, I introduce the virtual collaboration model stage by stage. The model is divided into four phases because its elements often distributed into four categories, as the data management levels (Kaizer and Kuiper 2014) and the social media use of scientists (I have left the last category out from the model, as it is not necessary to turn every scientist to a Cyberentrepeneur) (Nentwich and König 2012). Stage 1 of the model represents the current situation and is practically the summary of those outcomes of the interviews and the survey, which could be confirmed by literature. Stage 4 represents the ideal set of conditions that allows the use of VREs in their full potential. This was formulated from the literature review and the interviews with VRE developers. The stages in between can be understood as future recommendations for VRE developers, actual and prospective VRE user communities and research policy makers, and were made by using the literature review and the interviews with VRE developers. The detailed recommendations for distinct stakeholder groups can be found in the Discussions, in Chapter 0.

9.1.1 Stage 1

This phase represents more or less the current situation for life science researchers. From the technology perspective (Collaboration infrastructure readiness), VREs are in principle ready, but they are not perfect. To solve the problem of utility, collaborative co-design is suggested to be used, and several projects have this in their agendas. The co-design does not stop at that point when usually other projects stopped contact with users (after discussing their needs). User communities should be involved in the design and the definition of VREs, by listing those tools which are currently used by life scientist and integrate these into the infrastructure. Users also have to define how efficiently and easily can they use the current systems to help their further evolution. In terms of Collaboration technology readiness, the results of this research suggest, that the majority of life science researchers are at the basic level of data management, the majority use web1.0 communication tools, and they belong to the me-too-presence category in social networking tool use. Regarding collaboration readiness, at the moments institutions are formulating their data managements, privacy and safety policies. This is a significant step in helping scientists to move forward in terms of data stewardship.

9.1.2 Stage 2

From the technology perspective, at this point standardization (in terms of data management, linguistics and work practices) has to take place, to solve the problem of interoperability, funding issues, reliability and extendibility. In terms of Collaboration technology readiness, scientists need to change their data management habits and get used to the idea of data stewardship, metadata use, handling data privacy in place. The attitude of researchers towards social media should change so that the majority of investigators use social media at least as a digital calling card. Accreditation of activities outside the academic scientific publishing would increase the chances that they spend more time in engaging with communication ways and maybe with social media as well. Regarding collaboration readiness, institutional and research policy makers should focus on solving the problem of VRE funding and providing academic positions to maintain functions needed for the management and governance of VREs, to provide stability and sustainability of these infrastructures. The community aspects of the virtual research communities should get more attention.

9.1.3 Stage 3

Standardization introduced at Stage 2 should be implemented at higher levels and result in interoperability and compatibility of the different systems or VREs to minimize contra-productivity. With these changes and the use of co-design, VREs would be able to adjust constantly to the actual needs of various communities. In terms of Collaboration technology readiness, scientists need to step at a higher level in data management. Scientists should be encouraged to share their indexed data in searchable repositories, think in long terms regarding data management, use the data standards suggested by data management experts, and use high-level data interfaces. The role of institutes and research policy makers are crucial in this process. They also have to be encouraged to use social media as a networking tool, and get used to communicating with peers via these channels. These processes would also help the reinforcement of community feeling in virtual collaborations. The community uptake at this stage should focus on collaboration and research group leaders, as they have a more comprehensive view on collaborations, and see the added values of integrative infrastructures more. Moreover, in this research leaders were shown to be more active in using project management tools, academic social networking tools, doing science outreach.

9.1.4 Stage 4

Regarding data management, scientists should be helped to engage in flow-based systems, accept the use of infrastructure that integrates workflow and data management in platforms (VREs), and be connected to global communities. The attitude towards social media should further change, so that the majority of researchers use at least academia-related social networking sites actively, actively searching for potential networking partners, be involved in discussions and use other services of SNSs. By integrating postdocs and Ph.D. students into the community uptake process of virtual research communities, ideally, scientists would be able to adopt virtual research environments.

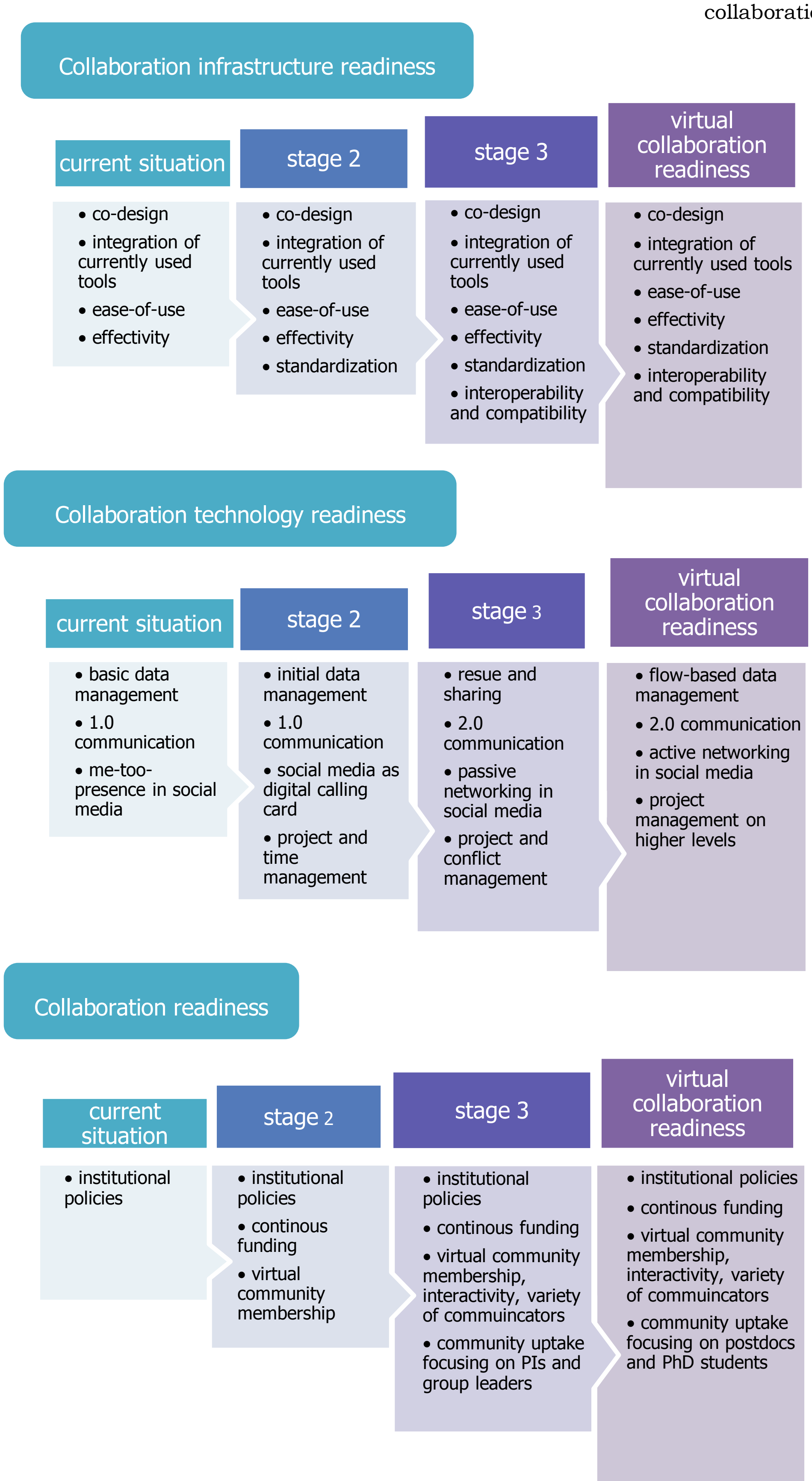

Figure 35. The virtual collaboration readiness model

10 Discussion

Virtual research environments are innovative, online, community-oriented, flexible, and secure working environments designed for scientific research groups working together (Candela, Castelli et al. 2013). VREs have the potential to change research practices, do the academic research faster, more efficient and transparent (Junge 2007), or even speed up the shift between fundamental research and applied sciences (Dutton and Jeffreys 2010). Grounded in these facts, it is not surprising that research policy makers at institutional or European levels would like to develop and employ virtual research environments (Dutton and Jeffreys 2010, EU retrieved at 28-04-2015). But scientists in life sciences think differently about this topic (Allan 2009). While the EU is investing into the development of monumental VRE projects and universities build their environments separated from each other, scientists would like to use their legacy systems. Based on research, scientists need access to data storage and computational resources (including grid computing), as well software and services, but they do not need VREs (Allan et al. 2006).

As the results of this research project show, most scientists in the field of life sciences are not aware of these virtual environments, so they cannot know the potential benefits of them. The lack of knowledge can strongly influence their attitude towards these highly advanced tools. In the research process, I collected information from different sources, from the literature and by interviewing VRE researchers, developer, and by interviewing and surveying potential and actual VRE users to build a model for virtual collaboration readiness. The basis of the theoretical framework used throughout the thesis was the technology acceptance model because virtual research environments are ICT tools that need to be accepted by their users. I assigned the three different factors of the technology acceptance model (namely the external variables, perceived ease of use and the perceived usefulness) to the three major dimensions of virtual research environments discussed by Olson et al as collaboration readiness: collaboration readiness, collaboration infrastructure readiness, and collaboration technology readiness (Olson, Teasley et al. 2002). I combined two other theories into the framework, and by investigating the results, I have built a stage-gate-like model to predict the virtual collaboration readiness of different collaborating groups.

At the first part of this chapter, I discuss the limitations of the research. In the second part of the discussion, I emphasize different aspects of the project, which are related to technology acceptance, communication, community aspects and to take an outlook to the large-scale VREs. Finally, I give guidelines and recommendations to diverse VRE stakeholders how to overcome the obstacles life science researchers face.

10.1 A critical view of the research

To cover as much factors as possible, I took a mixed approach to answering the research question. This made the argumentations of the study stronger, as most of the factors were mentioned in the literature, by the VRE developers and by life science researchers. The use of the mixed method could overcome some of the limitations of the research. The sample size I have used in my research was relatively small due to the limited resources I had. The interviews were performed with only eight researchers, and the survey was filled out by 76 scientists. This sample size represents 10% margin of error, at 90% confidence level and 0.5 standard deviations. This margin of error is considered relatively high. The sampling of the survey was also not random, as I reached out to scientists I knew, and I asked their help in distributing the link to the questionnaire. To overcome the errors arising from the sampling, I have incorporated only those results into the virtual collaboration model, which could be backed up by previous studies found in the literature.

The scope of this study was focusing on life science researchers; I have selected literature and designed the research with that focus in mind. Therefore, the factors I have selected are probably specific for this target group. Moreover, due to the fact that the survey was not validated prior to the research, it would not be useable for other studies. Based on these facts, the results of this study are not generalizable for other disciplines.

Due to the manual conversion of the answers given to the open questions of the questionnaire to Likert-scale-like answers, these data were not applicable to statistical analysis. The purpose of these conversions was to see whether there can be a trend observed between different practices and attitudes. Further research is needed to prove these trends.

Finally, the model was built on theories and results, but it needs to be tested to be improved. Further research is required to see the validity of the model by using life science researchers who stand at different stages of the models, and also to see if this can be applicable to other disciplines.

10.2 The adoption of VREs is affected by multiple factors

In the past two decades, VREs have been shown to speed up the research process by bridging steps in the research life cycle, to make research more efficient and expand the boundaries of knowledge generation and research methods (Junge 2007). Although these positive impacts would be desirable to scientists as well, some of these environments were unsuccessful (Bos, Zimmerman et al. 2007).

The acceptance of an infrastructure is a complex process and is affected by a variety of factors. To get a global picture of the interaction of these factors, different perspectives need to be combined. The majority of the literature focusing on VREs is written from the viewpoint of the technology (representing the infrastructure school of e-research), and investigate the applications and the details of the infrastructure itself, like utility, usability and so on (Bartling and Friesike 2014). If we take into account what the technology acceptance model is suggesting, namely that besides the perceived usefulness and perceived ease of use, external factors also affect the attitude of users, then we have to broaden our focus and look at aspects other than just the details of the infrastructure.

By involving communication (science outreach, social media use), project management and data management practices, community aspects and research policies, this research was trying to cover as many factors that can determine the attitude of life science researchers as possible. Throughout the thesis, I have used the perspective of the pragmatic school, which aims to involve external knowledge and helping collaborations with online tools. By involving science outreach into the research, I have taken the perspective of the public school as well (Bartling and Friesike 2014). In the interviews and the survey several other aspects of the e-research were mentioned, and I have incorporated some of them into the virtual collaboration readiness model as well. For example, I suggest to science policy makers that they extend the citation system by involving publications outside the classical academic journals, like in social media, or by incorporating data publishers to the citation system. These suggestions, as alternatives to the current impact factor system, belong to the focus of the measurement school of e-research.

10.3 Science Communication aspects of the research

Scientific work is a social process. Social interactions are present in all steps of the research life cycle in collaborations, from the development of new ideas to the execution of the research tasks. This gives the basis for applying the social network theory to scientific collaborations. The networks of collaborations can contain weak and strong interactions between the members. Communication is a fundamental part of scientific collaborations via connecting their members. Without proper communications, tasks could not be synchronised; scientists would not learn from each other, results would not be integrated, and it would be difficult to keep trust. The use of ICT tools can facilitate these communication tasks, especially in international collaboration (Sonnenwald, Solomon et al. 2002). In virtual teams, communication plays a central role to affect performance. Teams that function in a virtual environment, rely heavily on ICT tools, which causes more obstacles in information exchange (Qureshi, Min et al. 2005). The results of this research show, that many of the scientists rely heavily on personal communication. Some find video conferencing as a substitute for face-to-face communications while others don't, and they find personal interactions more efficient. Computer-mediated communication is also widely used by researchers, but this is dominated at the moment by sending information and data by emails.

A system which requires information exchange on the possible highest level has to fulfil the requirements of the communicating parties. While information is easy to store and transmit over long distances, **scientific knowledge**, which requires specialized expertise, is difficult to communicate (Szulanski, 1992). This knowledge changes quickly, may be implicit and difficult to represent or disseminate over big groups (Bos, Zimmerman et al. 2007). An infrastructure to support collaboration is suggested to work the best when it is completed with face-to-face communications (Hossain and Wigand 2004).

It may be true, that some scientific disciplines do not need VREs, only safe and reliable data stewardship, but if there will be an integrative infrastructure built to help collaborations, communication tools cannot be left out of the system. One of the bioinformatician PIs from the EpiPredict consortium summarized why it is important to investigate the communication, social networking, data management and project management practices of the potential users of VREs (interview #5, appendices, page 74). If it is not possible to integrate all the collaboration-related activities, which require a human-computer interaction, into the infrastructure, then the infrastructure will not be used. If scientists need to turn to a tool outside of the infrastructure to communicate, then the infrastructure will not be used.

10.3.1 Social media

In the process of designing the research and writing it up, I used the perspective of communication, more specifically science communication. I was interested in the communications practices utilized in a collaboration setup, in particular in computer-mediated communication means and social media. As virtual research environments include some aspects of social networking sites, the acceptance of these tools is particularly involved in the study.

Some of the most interesting and controversial results of this study were related to social media. While almost all scientists found networking necessary, they had very diverse opinions about social networking sites and their role in the academic world. Social media was thought to be dangerous by some life scientists, and generated intense feelings, like love, hate or annoyance. Most interestingly, roughly half of the scientists did not think about social media as a scientific communication form; it was declared as unscientific while 88% of the scientists filling out the VRE survey used social networking sites designed for scholars. It was also interesting to see that only 17% of the academic SNS-users were active users, the rest were just using these sites passively.

As the results of the VRE survey suggest a relationship between social media use and the acceptance of VREs (of course these results have to be confirmed by further research), it is worth to VRE developers to pay attention to what scientists think about social media, and potentially design communication plans to change their attitude if they want scientists to accept VREs. Moreover, as Bozeman and Corley suggested using scientists' social networks as a part of a collective measure, the scientific and technical human capital (S&T human capital), which is the combinations of scientific, technological and social knowledge and skills. It is built on all the formal education and training, social relations and network connections (Bozeman and Corley 2004). This is a new way to look at the competencies of scientists, by adding social networks to their portfolio. This approach can be linked to the measurement school of e-research, which focuses on extending the impact factors system with new ways of information sharing as public outreach or public engagement (Bartling and Friesike 2014).

10.3.2 Science outreach

Science outreach is one important aspect of the science communication agenda. **The public school of e-research** is focusing on this topic as well, namely the accessibility of the research process and the scientific results as well to a wider audience. Thanks to the new communication technologies, like the web2.0, scientists have much more opportunities to disseminate their findings in a comprehensive manner (Bartling and Friesike 2014). Based on this study and the results of the latest PEW thesis (2015), I thought that here should be a relationship between science outreach practices and social media use and that science outreach practices affect the acceptance of VREs, but these presumptions were not supported by the results of this research. Probably because science outreach is a broad term, and a variety of public engagement activities can be done without using social media.

10.4 EU-wide VREs

With the Horizon2020 einfra-9-2015 call the European Union has decided to invest into the development of virtual research environments. In the call description it was stated that "VREs should integrate resources across all layers of the e-infrastructure (networking, computing, data, software, user interfaces), should foster cross-disciplinary data interoperability and should provide functions allowing data citation and promoting data sharing and trust". Eight projects (EVER-EST, READ, MuG, OpenDreamKit, VRE4EIC, West-Life, BlueBRIDGE, VI-SEEM) were funded and launched following this call. The majority of them are Europe-wide projects.

On the one hand, all of the interviewed researchers were positive about the idea of building a huge EU-scale VRE to serve as a data repository and an SNS for scientists. They indicated that such a VRE could help keeping the data accessible, make science more democratic, transparent and increase efficiency, which is parallel to previous research results (Dutton and Jeffreys 2010). On the other hand, scientists filling out the survey had various thoughts about the usefulness of a VRE. 41% of the scientists were not sure whether VREs could be beneficial for scientific research while in total 57% found that virtual research environments could be useful. This is in parallel to the highly diverse attitude towards an EU-wide VRE. The major concerns about a huge infrastructure were related to the increase in bureaucracy, the size of the infrastructure (too big) and its efficiency. Data standards, data safety and security, legal and ethical issues related to patient data, equal access, and finally sharing data with competitors were also mentioned. One participant was against the concept because it would not be voluntary. From the positive side, the following arguments were mentioned: a place to publish negative results and detailed protocols, data would be freely available to the public, beneficial for collaborations. Some of these concerns and positive opinions can arise from the fact, that the majority of the scientists have not heard about VREs before, and were not aware of their functionalities and properties.

After the interview with Zhiming Zhao (appendices, page 46), it became apparent that technically there is no difference between small-scale and large-scale VREs, but there are some issues which could be different between the two. On the one hand, the sustainability of a big infrastructure could be more realistic if companies start to invest in the development of VREs. On the contrary, small-scale VREs could be more successful as those VREs were shown to be sustainable, which had an existing community behind these, and there are not so many pan-European scientific communities yet. Moreover, with the increase in the size of the community, communication will not be that open anymore, due to the growing amount of unknown participants. In a huge infrastructure, the fear from the competitors could also be an obstacle to sharing data (Sonnenwald, Solomon et al. 2002).

10.5 Community aspects of the research

Scientific collaborations are social interactions, and collaborating partners function in different levels in this complex social networks: on personal level, as individual researchers, with particular motives and goals; as individual users of the given ICT-tools and also on group level, as members of the collaboration (with distinct roles), members of a research group (with given positions), members of a research institute, members of a scientific discipline (with given working methods and jargon) and members of a cultural community (with given languages and cultural behaviour). The network theory is used throughout the whole thesis to provide an overview of the different levels of the collaborative network.

Virtual communities are *"webs of personal relationships in cyberspace"* (Rheingold 1993). If we put this term into research context, then we can define a virtual research community as a group of researchers working together and facilitated by a set of online tools, systems and processes interoperating to support collaborative research within a virtual research environment (Allan 2009). The results of the VRE survey showed that the highest percentage of scientists thinking about their collaborations as virtual research groups was in the group who participated in international collaborations. Moreover, those scientists, who considered their collaborations as virtual research communities, had a higher tendency to find virtual research environments useful for scientific research. This suggests that for those, who are interested in the adoption of VREs, it worth to concentrate on the development of virtual research communities. Based on Rheingold (1993) virtual groups are only called as virtual communities when they fulfil some criteria (Rheingold 1993). For remote international collaborations, this means that these form virtual research communities if these have a common public space, which could be provided by VREs, and which should contain communication tools that enable the minimum level of interactivity. If most members of the user community would use the infrastructure by sharing data, data-related information, and use communication tools to contact other members of the VRE user community, the variety of communicators would also be fulfilled. In the case of a small-scale, consortium-size VRE, the community feeling would be more realistic to build, while this could be probably problematic to achieve within a huge EU-wide VRE.

Another important community aspect of this research is that VREs can be different due to the various and potentially changing needs of the distinct user communities. User communities can be different due to the fact that scientists from different scientific fields were shown different tendencies in working together. The data they deal with during the research is also different, so the way they share and analyse these data is also different. Scientists from various disciplines have also found to have different priorities regarding the importance of the various elements of the research lifecycle (Allan 2009). Based on these differences several scholars suggested, that different disciplines would adopt VREs with various rates, due to the different cultural obstacles they face and due to various skills they need to learn to be able to use this infrastructure (Sonnenwald, Solomon et al. 2002, Allan 2009).

This could practically mean that different user communities may need different strategies to reach the virtual collaboration readiness state. Not every collaboration or virtual research community stands at the same level of the virtual collaboration readiness. Before the introduction of a VRE, the actual ICT tool use should be mapped in the given community, as suggested by Olson and his colleagues as well (Olson, Teasley et al. 2002). It is advised to map the ICT tools used by scientists in all of the following categories: communications, social networking, project management, workflow management, data management tools and other software used for research. If scientists are at an advanced level in one functionality, but not on the other, those areas should receive more attention that lag behind. It is also possible that some collaboration will never use collaboration-assisting tools because the scientists do not need most of the functionalities a VRE can provide. For example, sequential collaborations would only need solutions for data management, but international and integrative collaborations have a higher chance to use VREs in the future.

10.6 Guidelines and recommendations based on the virtual collaboration model

Taking the virtual collaboration readiness model as a basis, the list of things to be done to enable life science researchers to accept and use VREs is relatively long. The following guidelines are suggestions to several stakeholders to take actions if they would like to achieve the stage in which VREs could be used.

10.6.1 VRE developers

- First of all, as the definition of virtual research environment is relatively plastic, it would be advisable to find a consensus on what exactly VREs are, before the marketing of these infrastructures would begin.
- To solve the problem of utility, collaborative co-design is suggested to be used in the VRE projects. User communities should be involved in the design and the definition of VREs, by listing those tools which are currently used by life scientist and integrate these into the infrastructure. Users also have to define how efficiently and easily can they use the current systems to help their further evolution.
- VRE developers need to engage the user communities in this process and keep them involved in the co-design. To do that, it is advised to generate a strategy.
- Standardization of the of data stewardship processes, metadata use, but also the standardization of the jargon used in these infrastructure has to be defined to solve the problem of interoperability, funding issues, reliability and extendibility of VREs and to minimalize contra-productivity arising from the use of multiple systems.

10.6.2 Institutions, research policy makers

- As the majority of life science researchers are not aware of what the term VRE mean, after reaching a consensus on what the definition of VREs is, research policy makers should start to develop a strategy on how to reach scientist to inform them about VREs and to engage them in the co-design.
- Institutional and research policy makers should create a strategy to solve the problem of VRE funding and providing academic positions to maintain functions needed for the management and governance of VREs, to provide stability and sustainability of these infrastructures.
- Institutions need to formulate their data management, privacy and safety standards and policies and make the scientists aware of these regulations. This is a significant step in helping scientists to move forward in terms of data stewardship. Institutions should not forget to incorporate the financial part of these regulations into account.
- After defining these standards, research policy makers should design a motivating system that not only obliges but also encourages scientists to follow data stewardship protocols and to incorporate data-minded research planning into everyday practices.
- As intellectual property right, liability and copyright codes strictly regulate the choices scientists have to make in accepting and using a given infrastructure, these regulations should be modified to let scientists accept and use VREs.
- Parallel with this, it is advisable to accredit activities outside the academic scientific publishing, which would increase the chances that they spend more time in engaging with communication ways and maybe with social media as well.
- As group leaders have a more comprehensive view on collaborations, and as it was shown in this research, leaders were shown to be more active in using project management

tools, academic social networking tools, doing science outreach, institutions and research policy makers should first focus on collaboration and research group leader in the community uptake process, and only focus on postdocs and Ph.D students later.

10.6.3 Trainers, educators

- To help scientists change their data management habit and to help them get used to the idea of data stewardship, metadata use, handling data privacy in place, specific data management training and workshops should be developed.
- To change the attitude of life science researchers about social media, training and workshops should be made available to life science researchers.

References

Allan, R. (2009). Virtual Research Environments From portals to science gateways.

Bartling, S. and S. Friesike (2014). Opening Science, Springer.

Beukers, M. W. (2011). "Project management of life-science research projects: project characteristics, challenges and training needs." Drug Discovery Today **16**(3–4): 93-98.

Bonvin, A. M., A. Rosato and T. A. Wassenaar (2010). "The eNMR platform for structural biology." Journal of structural and functional genomics **11**(1): 1-8.

Bos, N., A. Zimmerman, J. Olson, J. Yew, J. Yerkie, E. Dahl and G. Olson (2007). "From Shared Databases to Communities of Practice: A Taxonomy of Collaboratories." Journal of Computer-Mediated Communication **12**(2): 652-672.

Bozeman, B. and E. Corley (2004). "Scientists' collaboration strategies: implications for scientific and technical human capital." Research Policy **33**(4): 599-616.

Bryman, A. (2015). Social research methods, Oxford university press.

Candela, L., D. Castelli and P. Pagano (2013). " Virtual Research Environments: An Overview and a Research Agenda." Data Science Journal **12**(0): pp.GRDI75–GRDI81.

Carusi, A. and T. Reimer (2010). "Virtual Research Environment, Collaborative landscape study." JISC-funded project.

Chen, Y., P. Martine, H. Schentz, B. Magagna, Z. Zhao, A. Hardisty, A. Preece and M. Atkinson (2013). "ENVRI Reference Model V1.1." http://www.envir.eu/en/rm.

Chompalov, I. (2014). Lessons Learned from the Study of Multi-organizational Collaborations in Science and Implications for the Role of the University in the 21st Century. The Institution of Science and the Science of Institutions, Springer**:** 167-184.

Chompalov, I., J. Genuth and W. Shrum (2002). "The organization of scientific collaborations." Research Policy **31**(5): 749-767.

Colen, R., I. Foster, R. Gatenby, M. E. Giger, R. Gillies, D. Gutman, M. Heller, R. Jain, A. Madabhushi and S. Madhavan (2014). "NCI workshop thesis: clinical and computational requirements for correlating imaging phenotypes with genomics signatures." Translational oncology **7**(5): 556-569.

Davis, F. D., R. P. Bagozzi and P. R. Warshaw (1989). "User acceptance of computer technology: a comparison of two theoretical models." Management science **35**(8): 982-1003.

De Roure, D., C. Goble and R. Stevens (2007). Designing the myexperiment virtual research environment for the social sharing of workflows. e-Science and Grid Computing, IEEE International Conference on, IEEE.

Dutton, W. H. and P. W. Jeffreys (2010). World wide research: Reshaping the sciences and humanities, MIT Press.

EU. (retrieved 07.03.2016.). "https://ec.europa.eu/digital-single-market/en/news/open-access-h2020-services-and-support-projects." Retrieved 08.03.2016.

EU. (retrieved at 28-04-2015). "http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/328-einfra-9-2015.html."

Fraser, M. (2005). "Virtual Research Environments: Overview and Activity " Ariadne **Issue 44**.

Goecks, J., A. Nekrutenko and J. Taylor (2010). "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences." Genome Biol **11**(8): R86.

Hara, N., P. Solomon, S. L. Kim and D. H. Sonnenwald (2003). "An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration." Journal of the American Society for Information science and Technology **54**(10): 952-965.

Hogue, T. (1993). "Community-based collaboration: Community wellness multiplied." Oregon Center for Community Leadership, Oregon State University.

Hossain, L. and R. T. Wigand (2004). " ICT Enabled Virtual Collaboration through Trust." Journal of Computer-Mediated Communication **10**.

Jankowski, N. W. (2010). E-Research: Transformation in scholarly practice, Routledge.

Jordan, K. (2014). Academics and their online networks: Exploring the role of academic social networking sites.

Junge, K. H., K.; Kelleher J. and Ramsden, R. (2007). "Formative Evaluation of the JISC VRE Programme

The VRE1 Programme: achievements and lessons learnt."

Kaizer, A. and T. Kuiper (2014). "IT for Research." TU Eindhoven.

Kraut, R., C. Egido and J. Galegher (1988). Patterns of contact and communication in scientific research collaboration. Proceedings of the 1988 ACM conference on Computer-supported cooperative work, ACM.

Lai, V. S. and H. Li (2005). "Technology acceptance model for internet banking: an invariance analysis." Information & management **42**(2): 373-386.

Lawrence, K. A. and N. Wilkins-Diehr (2012). Roadmaps, not blueprints: paving the way to science gateway success. Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond, ACM.

Ludäscher, B., I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. A. Lee, J. Tao and Y. Zhao (2006). "Scientific workflow management and the Kepler system." Concurrency and Computation: Practice and Experience **18**(10): 1039-1065.

Markauskaite, L., M. A. Kennan, J. Richardson, A. Aditomo and L. Hellmers (2012). "Investigating eResearch: Collaboration Practices and." Collaborative and Distributed E-Research: Innovations in Technologies, Strategies and Applications: Innovations in Technologies, Strategies and Applications: 1.

Mathys, T. (2004). The Go-Geo! Portal metadata initiatives. Proceedings of the Geographical Information Science Research UK 12th Annual Conference, University of East Anglia, Norwich, UK.

Moore, K. and P. Neely. (2011). "From social networks to collaboration networks the next evolution of social media for business ", retrieved at 10.03.2016.

Nentwich, M. and R. König (2012). Cyberscience 2.0: Research in the age of digital social networks, Campus Verlag.

Newman, M. E. (2001). "The structure of scientific collaboration networks." Proceedings of the National Academy of Sciences **98**(2): 404-409.

Olson, G. M., S. Teasley, M. J. Bietz and D. L. Cogburn (2002). Collaboratories to support distributed science: the example of international HIV/AIDS research. Proceedings of the 2002 annual research conference of the South African institute of computer scientists and information technologists on enablement through technology, South African Institute for Computer Scientists and Information Technologists.

Olson, G. M., A. Zimmerman and N. Bos (2008). Scientific collaboration on the Internet, The MIT Press.

Orchard, L. J., C. Fullwood, N. Galbraith and N. Morris (2014). "Individual differences as predictors of social networking." Journal of Computer ‐ Mediated Communication **19**(3): 388-402.

Pampel, H., P. Vierkant, F. Scholze, R. Bertelmann, M. Kindling, J. Klump, H.-J. Goebelbecker, J. Gundlach, P. Schirmbacher and U. Dierolf (2013). "Making Research Data Repositories Visible: The re3data.org Registry." PLoS ONE **8**(11): e78080.

Park, E., H. Kim and J. Y. Ohm (2015). "Understanding driver adoption of car navigation systems using the extended technology acceptance model." Behaviour & Information Technology **34**(7): 741-751.

PEW, r. C. (2015). How Scientists Engage the Public.

Piwowar, H. A., R. S. Day and D. B. Fridsma (2007). "Sharing detailed research data is associated with increased citation rate." PloS one **2**(3): e308.

Poole, M. S. (2009). "Collaboration, integration, and transformation: Directions for research on communication and information technologies." Journal of Computer ‐ Mediated Communication **14**(3): 758-763.

Porter, C. E. and N. Donthu (2006). "Using the technology acceptance model to explain how attitudes determine Internet usage: The role of perceived access barriers and demographics." Journal of business research **59**(9): 999-1007.

Procter, R., R. Williams, J. Stewart, M. Poschen, H. Snee, A. Voss and M. Asgari-Targhi (2010). "Adoption and use of Web 2.0 in scholarly communications." Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **368**(1926): 4039-4056.

Qin, J., F. W. Lancaster and B. Allen (1997). "Types and levels of collaboration in interdisciplinary research in the sciences." JASIS **48**(10): 893-916.

Qureshi, S., L. Min and D. Vogel (2005). A Grounded Theory Analysis of E-Collaboration Effects for Distributed Project Management. System Sciences, 2005. HICSS '05. Proceedings of the 38th Annual Hawaii International Conference on.

Rafaeli, S. and F. Sudweeks (1997). "Networked Interactivity." Journal of Computer-Mediated Communication **2**(4): 0-0.

Rheingold, H. (1993). " The virtual community: Homesteading on the electronic frontier." Reading , MA : Addison-Wesley. .

Ruzicka, L., Y. M. Bradford, K. Frazer, D. G. Howe, H. Paddock, S. Ramachandran, A. Singer, S. Toro, C. E. Van Slyke, A. E. Eagle, D. Fashena, P. Kalita, J. Knight, P. Mani, R. Martin, S. A. T. Moxon, C. Pich, K. Schaper, X. Shao and M. Westerfield (2015). "ZFIN, The zebrafish model organism database: Updates and new directions." genesis **53**(8): 498-509.

Savage, C. J. and A. J. Vickers (2009). "Empirical study of data sharing by authors publishing in PLoS journals." PloS one **4**(9): e7078.

Sciencegateways. "http://www.egi.eu/services/researchers/science-gateways/index.html."

Shackel, B. (1991). "Usability-context, framework, definition, design and evaluation." Human factors for informatics usability: 21-37.

Sonnenwald, D. H. (2007). "Scientific collaboration." Annual Review of Information Science and Technology **41**(1): 643-681.

Sonnenwald, D. H., P. Solomon, N. Hara, R. Bolliger and T. Cox (2002). "Collaboration in the large: Using video conferencing to facilitate large group interaction." Knowledge and information technology in 21st century organizations: Human and social perspectives: 115-136.

SURF (2011). VRE Starters Kit. information retrieved from the SURF website.

Teif, V. B. (2013). "Science 3.0: Corrections to the Science 2.0 paradigm." arXiv preprint arXiv:1301.2522.

Thagard, P. (1997). "Collaborative knowledge." Noûs **31**(2): 242-261.

van der Vaart, L. (2010). "Collaboratories: connecting researchers: how to facilitate choice, design and uptake of online research collaborations.".

Van Eperen, L. and F. M. Marincola (2011). "How scientists use social media to communicate their research." J Transl Med **9**: 199.

Van Noorden, R. (2014). "Online collaboration: Scientists and the social network." Nature **512**(7513): 126-129.

Walsh, J. P. and T. Bayma (1996). "Computer networks and scientific work." Social studies of Science **26**(3): 661-703.

Walsh, J. P. and N. G. Maloney (2007). "Collaboration structure, communication media, and problems in scientific work teams." Journal of Computer - Mediated Communication **12**(2): 712-732.

Wray, K. B. (2002). "The epistemic significance of collaborative research." Philosophy of Science **69**(1): 150-168.

Wulf, W. A. (1993). "The collaboratory opportunity." Science **261**(5123): 854-855.