



universität
wien

Final Report

As part of the research practice
Social and environmental sustainability - Trade-offs and opportunities

„How to make research more sustainable“

Submitted by
Clemens Posselt

Vienna 2022

Supervisor:

Univ.-Prof. Yuri Albert Kyrill Kazepov, PhD
Tatjana Boczy, BA MA

Table of content

1.	FIRST SEMESTER.....	3
1.1	INTRODUCTION.....	3
1.2	SYSTEMATIC LITERATURE REVIEW.....	4
1.2.1	<i>Definition of Sustainability.....</i>	4
1.2.2	<i>Summary of the literature.....</i>	5
1.2.3	<i>Differences and commonalities.....</i>	7
1.3	CASE STUDY.....	8
1.4	ANALYSIS & DISCUSSION.....	10
1.5	CONCLUSION.....	11
2.	SECOND SEMESTER.....	11
2.1	INTRODUCTION.....	11
2.1	CURRENT FRAMEWORK.....	12
2.1.1	<i>Repositories.....</i>	12
2.1.2	<i>Metadata.....</i>	13
2.1.3	<i>Guidelines of universities.....</i>	13
2.2	FIRST INTERVIEW.....	14
2.2.1	<i>Status Quo at WU.....</i>	14
2.2.2	<i>Different kinds of repositories.....</i>	15
2.2.3	<i>Responsibility of the researchers.....</i>	16
2.2.4	<i>Open Access at WU.....</i>	16
2.2.5	<i>Re-use of data.....</i>	17
2.3	SECOND INTERVIEW.....	18
2.3.2	<i>Status Quo of data sharing.....</i>	18
2.3.3	<i>PHAIDRA.....</i>	19
2.3.4	<i>Responsibility of the researchers.....</i>	20
2.4	DISCUSSION AND POLICY RECOMMENDATIONS.....	20
3.	REFERENCES.....	23
4.	APPENDIX.....	26
4.1	TRANSCRIPT INTERVIEW 1.....	26
4.2	TRANSCRIPT INTERVIEWEE 2.....	34

1. First Semester

1.1 Introduction

In the course “Research Practice 1: Social and environmental sustainability - Trade-offs and opportunities”, all participants will perform different tasks to make their own research on a specific topic. The aim is to experience what it is like to work full-time in academia by performing all tasks. First, it is essential to analyze the literature that has been done before, as in modern days, you can use the knowledge that has been created by other scientists who did already work in this field. Then the idea is to use the literature as basis for the own research.

At the start of the course the participants were assigned to four different groups with the following topics:

- Infrastructure
- Organizational culture
- Research
- Third Mission

The topic I am supposed to deal with is research. I am supposed to find out what possibilities there are to make the research that is being done socially and ecologically sustainable. To do this, I will first conduct a systematic literature review, meaning that I will look at already existing literature by different authors meaning that I will analyze the different results, research questions and theories of the various papers and relate them to each other in order to identify differences and commonalities. Based on this, a case study will be planned and carried out to apply the knowledge gained from the systematic literature review to Austrian universities.

1.2 Systematic Literature Review

At the start of the subject “Research Practice 1: Social and environmental sustainability - Trade-offs and opportunities”, the systematic literature review was performed to build a basis for further tasks. The topic the author of this report was assigned to was “Research”, and as a result the search for literature that dealt with sustainable research had to be done.

In order to find as many relevant texts as possible that deal with this topic, I used the software “Web of Science”. First, the search string "SUSTAINABLE RESEACH" was used and received more than 430.000 search results. 430.000 were too many results and much of the content shown was not really relevant to the question. Therefore, the search was limited by a lot, by limiting the results to the following Web of Science categories: “SOCIAL SCIENCES INTERDISCIPLINARY” and “SOCIOLOGY”. This already had a major positive impact on the results, and only 6.400 results were displayed. Nevertheless, this number was still too high and should be reduced. When looking at the results, it is obvious that many do not deal with the topic of "sustainable research" but with “research on sustainability”. That is why the search sting ""SUSTAINABLE RESEARCH"" in inverted commas was used to get more relevant results.

This was also the final setting, 364 results were shown and on top a good proportion of which were relevant results. By only showing the category "SOCIAL SCIENCES INTERDISCIPLINARY" the results were narrowed down to 12 results.

1.2.1 Definition of Sustainability

After reading the literature, it was clear very soon, that there is not one definition of sustainable research alone, and because of that there are many papers which address different topics but with the same name: sustainable research. One can roughly divide the different texts into two groups

1.2.1.1 Sustainability: Long-term research infrastructure

One group is understanding “sustainable research” as a long-term possibility for some regions, for example developing countries, to conduct research (Gomo et al., 2011; Mashaah et al., 2014). Some regions in the world do not have access to infrastructure for research and therefore research is not really possible in the long-term. Authors addressing this topic are suggesting possibilities to solve this problem to enable research in a sustainable way in these countries. Not only in developing countries but also in developed countries, there are issues. Ankur Chandra (2013) for example writes about the cut of fundings, which can hinder research. Money is important for research and if there is less available, a lot of research cannot be done. Based on the author, this can be counteracted with transdisciplinary cooperations.

1.2.1.2 Sustainability: Resource-efficient research

The second group of texts has a different understanding with the term “sustainable research” and connects it with other ideas. In summary, it is about using resources more efficiently to make research more sustainable. Since this understanding of the term “sustainable” is more suitable in the context of the subject “Research Practice 1”, this group of authors is more relevant to analyze in more detail than the first one. This is because the goal of this subject is to create a project for the University of Vienna, which is not based in a developing country and is not yet much struggling with budget. First, the texts of the second group will be analysed and then the different ideas to face the problems will be compared to get an idea about the case study, which will be performed in this semester.

1.2.2 Summary of the literature

The paper of Ausina and Maté, written in 2010, is about primary data in Psychology. The authors criticize that primary data is rarely shared in Psychology and in their opinion this practice is not sustainable. Based on the authors, the main problem is that primary data is not shared, although there are enough possibilities. They see the reason in a lack of tradition and personal skepticism when being asked to share data, second also connected to effort.

Ausina and Maté (2010) recommend new ways to share primary data without effort and also recommend strategic alliances between researchers.

Akhlesh Lakhtakia (2010) criticizes that many resources are wasted by conducting a lot of research more than once, even though this is unnecessary. He suggests that international work should be done to take into account research from all over the world. He also suggests holding small conferences as webinars to make them more accessible. Overall, his approach is to make research more international to make it more sustainable.

The paper of Gottfried Tappeiner, Ulrike Tappeiner and Janette Walde (2007) suggests a similar idea. In their opinion, the increasing complexity of research questions as well as answers leads to the need for transdisciplinary cooperation in order to answer these questions in a sustainable way. To ensure this transdisciplinary approach, the authors write:

“Embedding various research projects in a more global framework can meet the demands of an interdisciplinary or transdisciplinary approach (especially for early stage researchers) and contribute significantly to sustainable research.” (Tappeiner et al., 2010, S. 253)

Pierre Tremouilhac et al. (2020) dealt with the problem of too little shared data. The problem of insufficient data in his local case was solved by programming a storage location where chemical data can be stored and shared. In the opinion of the author, this could lead to more sustainable research.

Mehmed Belgin et al. (2019) address the issue of sustainable data storage with the aim of ensuring that all data remains stored in the long term and does not disappear due to obsolete systems. They describe “a sustainable strategy to support a large number of researchers with widely varying scientific software needs, which is a common problem for most centralized Research Computing Centers on university campuses” (Belgin et al., 2019, Summary).

1.2.3 Differences and commonalities

All in all, it can be said that the topic of "sustainable research" is very complex and has many facets. However, it is visible that the aims of the authors just mentioned are the same. They all want science to use resources efficiently by also considering already existing research. They all criticize that this is not the case yet and that existing knowledge is not taken into account. It seems that all the authors are pursuing the same goal and complement each other. While Ausina and Maté (2010), Belgin et al. (2019) as well as Tremouilhac et al. (2020) have dealt with local solutions for data sharing, Tappeiner et al. (2007) and Lakhtakia (2010) have a global approach.

All authors are addressing the same problems and have similar approaches to solving them. The issue is that too little data is shared and therefore a lot of research is replicated, which is not bad per se, but wastes a lot of resources. This is especially the case with primary data, which can be analyzed multiple times with different techniques.

To solve this, all authors suggest sharing the data as easily as possible via software. Furthermore, not only researchers from the same discipline should be able to access the data, but also researchers from other disciplines. While this was particularly emphasized by Tappeiner et al. (2007), the other authors would probably also agree with this, at least there is no contrary opinion that has been written down.

In conclusion, it should be emphasized once again that there is basically a consensus among the scientists analyzed, even if they do not have the same proposal for solving problems in this area. However, the solutions are very complementary and it suggests that international, interdisciplinary and efficient data storage is the ultimate goal of all the researchers. Storing and finding data should be as simple as possible, as recommended by Ausina and Maté (2010), and should require little effort.

1.3 Case Study

After the SLR was completed, a case study was conducted with the idea to put the knowledge that was gained into context using the practical example of the University of Vienna.

In the course of this, it was found out about the databases of the University of Vienna and a closer look at them was taken. It was noticed that no universal database is provided by the institution. Access to many different international databases is granted, but there is no single universal database. The extent of the many different databases was overwhelming which can be shown by the fact that the University of Vienna has 58 different databases for the Department of Sociology alone.

However, the University of Vienna has the tool u:search, which you can use even if you do not confirm your identity, so even if you are not a student at the University of Vienna, you can use this tool. The use of u:search is described as follows:

“Search in print and electronic publications (books, journals and articles) in the holdings of the UB Vienna as well as in licensed or freely accessible databases and e-journal collections” (Universität Wien (b), 2021).

As can be seen, only literature from licensed databases can be found. Databases for which the University of Vienna does not have a license are therefore not displayed. Nevertheless, the literature that can be found is very extensive, as can be seen from the fact that the search term "Sociology" yields more than four million results, of which more than half (2.1 million) are articles from peer-reviewed journals. International papers are also included, you can find scientific contributions in 70 languages if you search for "Sociology" (Universität Wien (b), 2021).

However, limitations become apparent as soon as one restricts the search to "Open Access". Then only about 730,000 search results are displayed, for all other results you need licenses,

have to pay money or have to overcome other obstacles. Less than one fifth of the sociology literature found by u:search is freely available.

It is difficult to find out how large the share of transdisciplinary research is, because the classification of individual papers on specific disciplines and topics is not transparent. You can limit the results of your search and select topics in which you are interested but it is not possible to see the classification of each research.

Apart from the u:search tool, the University of Vienna also offers AUSSDA (AUSSDA, 2022). AUSSDA stands for "The Austrian Social Science Data Archive" and is a digital infrastructure for archiving data and at the same time functions as an aid to finding secondary data. Since June 2020, AUSSDA has been certified with the Core Trust Seal (CTS) as a "trustworthy data repository" (AUSSDA, 2022). One advantage over other platforms is that relevant data can be obtained within a very short time. With comparable platforms like Eurostat, the formal process takes several weeks or even months until you get what you are looking for.

In the Institute of Sociology, there is also the so-called "Datencenter" (2022), which is also intended to facilitate the search for secondary data.

Besides analyzing the databases of the University of Vienna, also a look into "Going Green", a sustainability-concept created by "Arbeitsgruppe für Nachhaltigkeit am Institut für Soziologie", short AGNIS (2021), was taken. The working group presents various measures to make the University of Vienna more sustainable. The four topics to which were also assigned in this course were given special consideration. Accordingly, a closely look at the proposed measures that were assigned to the area of "Research" was taken.

In "Going Green" (2021), the promotion of interdisciplinarity and transdisciplinary is recommended. Measures aimed at using resources more efficiently were not elaborated.

In order to evaluate how good the tools provided by the University of Vienna are, the possibilities provided by the Vienna University of Economics were compared with them.

When comparing the databases that the universities have access to, there is a clear difference. The University of Vienna has included more databases for sociology alone than the WU provides in total.

Furthermore, the WU does not provide tools such as AUSSDA or the "Datencenter" to facilitate the use of secondary literature.

1.4 Analysis & Discussion

As described in the Case Study, the University provides a way to easily search for existing literature. If you are a student at the University of Vienna or otherwise employed, you can use a large part of the literature you have found. Overall, u:search is a system that includes many scientific papers of various topics and thus represents a universal tool for literature search, which was suggested by Tremouilhac et al. (2020).

In addition, u:search also takes into account international and interdisciplinary work, which is in line with wishes and requirements of Lakhtakia (2010) and Tappeiner et al. (2007). One demand of the two authors, which was not taken into account by the recommendations of AGNIS (2021), is that research projects should be planned in an interdisciplinary and international way in the future. This measure is not included in "Going Green".

The University of Vienna also does not meet the suggestions of Ausina and Maté (2010), who suggest an initiative that creates an incentive to share primary data. This system should also be easy to use and should not involve a lot of effort.

What was not able to explore in the case study was the idea of Belgin et al. (2019) who said that a database should meet the needs of all disciplines included. For this, interviews with representatives of the individual disciplines would have to be done to gain insight into whether u:search meets the requirements.

1.5 Conclusion

In the analysis of the existing literature, many suggestions were found that could lead to the efficient use of resources. The research of the case study showed that parts of these suggestions have already been implemented in the University of Vienna, because a suitable tool, u:search, exists to find already existing literature. However, it is unclear whether it is efficient or still needs improvements to unfold its full potential.

Measures apart from the efficient use of existing data are the promotion of interdisciplinary and international research, which is partly fulfilled by the University of Vienna. Although the University of Vienna cooperates with international researchers and has a field of many disciplines, any concrete promotion in this direction were not particularly seen.

These missing items should be the focus of the second semester. Additionally will the efforts of the University of Vienna to promote interdisciplinary and international research examined. At the end, this requirement is central for more sustainable research based on most of the authors analysed by the systematic literature review

2. Second Semester

2.1 Introduction

In the second semester, the task was to formulate concrete measures to make research at the University of Vienna more sustainable. For this purpose, the results of the first half of the course were analysed again and it made the most sense to focus on the efficient use of resources. The reason for this is the enormous number of scientific documents and data that are not used, as also criticised by Ausina and Maté (2010). A lot of data is not shared or not used despite sharing. The aim of the second semester is to go more in depth into this topic and also to find out what concrete obstacles exist in reality.

In order to answer these questions, two interviews were conducted during the second half of the seminar to find out the opinion of experts on the topic of "sustainable research". One expert is an important person in relation to the data repository PHEIDRA, which is located at the University of Vienna, and the other person is a data manager at the Vienna University of Economics, one of the largest universities in the city. The aim of these semi-structured interviews was to discuss the challenges of sustainable research with people who deal with them on a daily basis in order to generate new insights.

2.1 Current Framework

Before conducting the interviews, the current framework conditions in connection with the sharing of data were analysed. The first step was to research which data repositories are available at the University of Vienna and can be used by researchers at the university.

2.1.1 Repositories

AUSSDA and PHEIDRA are the repositories of the University of Vienna. AUSSDA offers an infrastructure for social science data, also with a focus on ensuring that the data continues to be used after a project. (AUSSDA, 2022). It is certified as a "trustworthy data repository". There are 4,349 objects on AUSSDA, of which 1,289 are datasets.

PHAIDRA is also a repository, but it is not specialised in social sciences. The aim of PHAIDRA is to permanently secure and archive digital holdings at the University of Vienna (PHAIDRA, o.d.). As of 12 July 2022, PHAIDRA has recorded 6553 objects, of which all objects have unrestricted access, i.e. the meta-description of all objects can be viewed. The terms of use depend on various licences.

There are of course other, international repositories. A large part of them can be found via re3data (Re3data, 2022). In total, 2908 different repositories can be found on re3data, 307 of which are repositories for social sciences. AUSSDA can be found via re3data, but PHAIDRA cannot.

2.1.2 Metadata

In the course of the discussion about repositories, it is important to define the term metadata.

"Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information." (National Information Standards Organization, 2004).

The better this metadata is defined and described, the easier it is for researchers to find relevant data sets. In addition, for the person who collected the data, it has the advantage that the data may be used more often because its visibility is increased.

2.1.3 Guidelines of universities

Apart from the existing repositories, the university guidelines had to be found out. For data to be reused, it must first and foremost be available. Open Access means that everyone has access. The University of Vienna supports researchers in making their publications available online. To this end, there is an extensive funding program that enables university members to publish in more than 14,000 journals in Open Access (Universität Wien, o.D (c)) . However, Open Access is not mandatory.

In order to analyse another university as well and to compare it with the University of Vienna, the University of Economics was also analysed. WU also sees Open Access as an important goal:

"Die WU sieht es als ihre gesellschaftliche Verantwortung an, einen unbeschränkten und dauerhaften Zugang zu den wissenschaftlichen Publikationen ihrer Angehörigen zu gewährleisten. Daher hat sie sich zum Ziel gesetzt, an der WU entstandene Forschungsergebnisse nach dem Prinzip von „Open Access“ möglichst umfassend frei über das Internet zugänglich zu machen." (Wirtschaftsuniversität Wien, 2016)

However, WU's policy is only a recommendation; the open access provision of data or publications is not obligatory. However, the aim is for every publication by WU Vienna researchers to be published in the university's own repository "ePub", with open access.

2.2 First interview

In the following, the central results of the interviews that were conducted are presented. First, the statements of the first interviewee are shown, who, in addition to research data management at the University of Economics (WU), also deals with literature and information procurement in the field of mathematics, statistics and economics.

2.2.1 Status Quo at WU

According to this person, the Austrian Ministry of Science had asked the universities to deal with the topic of data management and to develop a strategy. It was this request that brought him into contact with research data management so intensively for the first time. By now, there are data management policies at all the major universities in Vienna. Austria, however, is clearly a latecomer. In the USA, this has been an important topic since the 1960s, which is related to genome and medical research, according to the interviewee. There, data was shared via repositories at an early stage.

The person emphasises that different cultures prevail in different departments. This begins, for example, with the data itself. Lawyers, for example, would hardly produce data comparable to the social sciences. The latter often have a qualitative methodological approach, which makes sharing this data difficult for data protection reasons.

At WU, they try to make research more sustainable by providing information on this topic to the employees. In addition, it is now standard at WU to create a data management plan when carrying out a project for the university. In the case of the Vienna University of Economics, this is made more difficult by the fact that it does not have its own data repository, so the researchers have to provide the data externally. For this purpose, the following repositories are accessed, for example: AUSSDA, a repository of the University of

Vienna, and Zenodo. Zenodo is a generic repository where you get a digital identifier (DOI) and can use it to reference the dataset.

At the University of Economics there is ePub, a digital storage platform for publications. Some of the ePub metadata can also be found in international portals, but not every metadata.

2.2.2 Different kinds of repositories

When asked why there are so many different repositories, the person replies that there are roughly the following three different forms of repositories.

2.2.2.1 Generic repositories

Generic repositories offer the possibility of uploading any data, regardless of its nature, and storing it long-term. Zenodo is one of the largest generic repositories. You can upload images, audio files, data files or many other documents.

2.2.2.2 Subject-specific repositories

Compared to other repositories, subject-specific repositories have the advantage that they are created specifically for certain data structures. This means that there are no technical problems. Furthermore, the metadata description is standardised and therefore it is easier to find relevant literature.

2.2.2.3 Institutional repositories

Institutional repositories have the advantage that researchers from the respective institution can upload their data quickly and easily. In addition, the institution has a certain data sovereignty, which is a political wish. If you fund research, then you want to have a say in it and do not want to have the data stored in a generic repository. According to the interviewee, this form of repository is most likely to disappear.

According to the interviewee, researchers have preferred to store data in subject-specific repositories until now, as these are often adapted to the respective data structure.

2.2.3 Responsibility of the researchers

Researchers at the University of Economics can ask for support at the so-called research service, but there is no central department. There are not enough resources for this and it is also expensive for the university. For this reason, researchers have to take care of these issues themselves.

In regards to a data management plan, the interviewee said that even though a data management plan is obligatory for FWF funding, this does not conversely mean that every research project has a data management plan. For example, about 80% of research at the Vienna University of Economics is not funded. For this reason, WU-internal projects have also been required to prepare the data management plan since 2017.

The claim that the FWF requires a data management plan turned out to be true in subsequent research. The purpose of the "Wissenschaftsfonds zur Förderung der wissenschaftlichen Forschung" (FWF) is to allow the sciences to develop at a high international level by providing financial support for scientific projects (Wissenschaftsfonds zur Förderung der wissenschaftlichen Forschung, o.D. (b)). In 2021, 732 applications were approved and funded with an average of 354,534 euros. This leads to a total amount of funding of 259,518,888 euros (Wissenschaftsfonds zur Förderung der wissenschaftlichen Forschung, o.D. (a)) . The influence of the FWF is evident from this large sum.

The FWF's Open Access Policy states: „In diesem Sinn verpflichtet und fördert der FWF alle ProjektleiterInnen und ProjektmitarbeiterInnen, ihre referierten Forschungsergebnisse im Internet frei zugänglich zu machen, wenn diese vollständig oder teilweise aus vom FWF geförderten Projekten hervorgegangen sind“ (Wissenschaftsfonds zur Förderung der wissenschaftlichen Forschung, o.D. (c))

2.2.4 Open Access at WU

The interviewee also stated what the current status but also the difficulties of Open Access are at the University of Economics. There is an OA policy at WU that recommends publishing

as openly as possible. In addition, according to the interviewee, the best journals themselves also advocate an OA culture. This is for various reasons and is intended to ensure reproducibility, especially in quantitative research. In addition, the collection of primary data is very expensive, so the data should be re-used with another research question. In the case of data that are not suitable for alternative research questions, there is no point in making them public.

Moreover, the usefulness of OA also depends on the discipline. Sometimes this publication of data is too costly. If OA does not bring added value for researchers, then the cost is too high and the benefit too low.

In addition, different disciplines have different value systems. In mathematics, for example, it is common practice to share data, whereas in other fields it is different. If a researcher can make further publications with the data he has collected, then the motivation to share this data is low. Ultimately, there is a lot of competition in science. In science, you live by publishing in good journals, your career depends on it.

2.2.5 Re-use of data

Lastly, the topic was the re-use of data. The interviewee stated: „Bei der Suche ist das Problem, dass man nicht weiß ob es den Datensatz bereits gibt. Ein zweites Problem ist die Verknüpfung von Daten, welche frei zugänglich sind in Kombination mit Daten, die man lizenzieren muss. Dann habe ich Daten die aus zwei Teilen bestehen und dann habe ich das Dilemma, wenn ich es Publizieren will, dass ich das nicht kann“ (Interviewee 1, p. 33). This dilemma is another argument for mandatory open access publication of data. Because only OA makes it as easy as possible for researchers to work with data and leads to more sustainable research.

„Ich glaube, dass Forscher wissen, wo sie bestimmte Daten suchen müssen, weil sie im regen Austausch untereinander stehen. Die Suche nach einem Datensatz ist schwierig, wenn man nicht weiß wie sie beschrieben sind. Man braucht jemanden, der weiß, wie die

Daten beschrieben sind. Die Suchstrategie ist individuell, viele googlen einfach, aber einen Datensatz findet man so eher nicht. Aber es gibt einige gute Indexierungen, z.B Base.“ (Interviewee 1, p. 33).

In addition, one can search meta-repositories for repositories that contain the topic the researcher is interested in. In this way, one can better narrow down the search. Good indexing is essential for this. According to the interviewee, there are no central standards for metadata at the University of Economics. There are various standards, such as the Dublin Core, but their use is not obligatory everywhere. Therefore, the metadata descriptions differ depending on the repository.

The interviewee stated that a central search engine for data that would find relevant data sets using an algorithm would of course be ideal. However, this technology does not exist at the moment.

2.3 Second interview

The second interview was conducted with a person who helped set up the University of Vienna's repository, PHAIDRA. Building on this, she is involved in research data management, data management plan consulting and similar topics.

2.3.2 Status Quo of data sharing

In her opinion, the topic is becoming increasingly important. On the one hand, from the university side and on the other hand, also from the funding bodies. She refers to the FWF, which since 2019 has required a data management plan from all researchers who receive funding. According to the interviewee, it used to be the case that no one was interested in the data as soon as a project was published. Now, in every discipline, you have to specify what is done with the data you generate and use. This is becoming more and more important for various reasons. On the one hand, one wants to save money, after all, data collection costs a lot of money; on the other hand, one wants to make the process more

transparent and comprehensible. Because data was rarely shared in the past, there would be weaknesses in this area today. This is also shown in the following research.

In 2018, only 14% of researchers published their data in a data repository. A much higher proportion of 27% published the data separately and 35% appended the data to the associated publication (European Commission, 2019). This shows the high catch-up demand of researchers to share their data in repositories. The advantage of this would be that the data is easier to find through the metadata description and could also be used more easily for further research, leading to more sustainable research in general.

There had once been efforts in Austria to set up a national repository, but this was rejected. Every university has tried to set up its own repository or to buy into external ones.

2.3.3 PHAIDRA

PHAIDRA has 20 partners who access the system. The goal of PHAIDRA is that it should be an archiving place for all kinds of objects, whether pictures, documents or sound recordings. Compared to AUSSDA, however, the data is not checked for quality and even the metadata has to be formulated in PHAIDRA itself. The reason why this is not done is the cost. The necessary personnel capacities are not available.

Most of the documents on PHAIDRA are open access. Nevertheless, researchers have the possibility to block their documents for external persons or to use licences so that they can only be used for certain purposes. But even beyond arbitrary decisions by researchers, there would be further obstacles to making OA data available. Legal frameworks play an important role, especially for qualitative methods. For example, the declaration of consent from interview partners is obligatory for recording, transcribing and publishing interviews. However, this declaration can also be revoked, and many researchers do not want to face the possibility of the extra work involved in removing this interview from many possible repositories. In the past, this topic was not of great importance, so many researchers do not yet have any contact with this topic. According to the interviewee, there is a process

underway for researchers to become more aware of legal issues. Some are already advanced in this while others are still at the beginning. The person sees a similar process in the fact that researchers have to learn to find out what already exists. For this, searching and finding data is essential, the person's opinion is that this is a skill that needs to be learned.

2.3.4 Responsibility of the researchers

In the course of the metadata description, the researchers themselves are responsible for describing the metadata as accurately as possible. They receive support, but PHAIDRA does not actively intervene in the process. Currently, however, the PHAIDRA user interface is being revised, in which case there might be a different metadata scheme.

About the responsibility of the researchers, the interviewee said: „(...) grundsätzlich ist es so, dass ForscherInnen aufgefordert werden so einen Managementplan abzugeben, aber viele haben sich damit noch gar nicht auseinandergesetzt. Es gibt viele Fragen die vor 10 Jahren noch nie ein Thema waren (...)“ (Interviewee 2, p. 35)

2.4 Discussion and policy recommendations

Certain overlapping opinions emerged in both interviews. It is clear that the efficient use of data is still far from the level it should be. However, there is an intention both at WU and from PHAIDRA to push this issue and enrich science so that data is used more sustainably in the future.

Nevertheless, there are some problems. The main problem is that researchers have to deal with all the issues on their own, even though many of the aspects were not part of their training. Also, researchers who have been publishing for a long time may have problems changing their existing habits. If they haven't used the relevant repositories for 20 years, how are they supposed to know which repositories to use? The fact that researchers are currently responsible for this is a current structure that will not be easy to change. It is time that researchers are supported when it comes to the issue of sustainable research.

Researchers currently have to create a data management plan on their own before their project, search repositories to find relevant data. After the project, they have to deal with where to publish themselves and researchers also have to work out the metadata of the collected data or the publication on their own.

Researchers should receive information on where best to publish their own data or work. There are many options, but they differ in quality and reputation. So that each researcher does not have to make up his or her own mind about all the options, which takes a lot of time, this information should already be available. In addition, there are 2,908 different repositories on re3data alone. Sifting through them all is an impossibility. Therefore, high-quality repositories should also be part of the basic knowledge of researchers. Furthermore, researchers have to create a data management plan for most projects. The FWF provides a template, and the institute should do the same for their staff.

Finally, the topic of metadata should become a stronger focus of the institute. Metadata is essential to find data and to be found. When collecting data, it must be adequately described with metadata. In addition, there should be contact details in case researchers need support in a certain area. They should not waste their time looking for contact details, that is not sustainable.

In summary, I recommend the institute of Sociology at the University of Vienna to provide a handbook with clear information on the following topics mentioned:

1. Where can data/papers be published?
2. Relevant repositories for each method/topic, etc.
3. A standardised open access data management plan
4. Contact information for experts on various topics (legal, data management...)
5. Guidelines for creating optimised metadata descriptions to be found

With this measure, all researchers at the Institute are helped. Those who have already mastered all these aspects perfectly have no disadvantage with the manual and all others

benefit from it. In order to keep the manual up to date, it should be checked once a year for correctness. In this way, research at the Institute of Sociology can be made more sustainable.

3. References

AGNIS. (2021). Going Green.

AUSSDA. (2022). AUSSDA. <https://aussda.at>

Ausina, J. B., & Maté, C. O. (2010). Data sharing: Towards a more sustainable research. *Psicothema*, 22(2), 263-269.

Belgin, M., Perini, T. A., Liu, F. C., Zhang, N., Sarajlic, S., McNeill, A., Manno, P. & Bright, N. C. (2019). A data-driven support strategy for a sustainable research software repository. *Concurrency and Computation: Practice and Experience*, 31(20). <https://doi.org/10.1002/cpe.5338>

Chandra, A. (2013). Multidisciplinary collaboration as a sustainable research model for device development. *Journal of vascular surgery*, 57(2), 576-582.

Datencenter. (2022). Datencenter. <https://www.soz.univie.ac.at/forschung/datencenter/>.

European Commission. (2019). Facts and Figures for open research data. https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor/facts-and-figures-open-research-data_en

Gomo, E., Kalilani, L., Mwapasa, V., Trigu, C., Phiri, K., Schmidt, J., & van Hensbroek, M. B. (2011). Towards Sustainable Research Capacity Development and Research Ownership for Academic Institutes in Developing Countries: The Malawian Research Support Centre Model. *Journal of Research Administration*, 42(1), 38-45.

Lakhtakia, A. (2010). Sustainability research and sustainable research. *Journal of Nano photonics*, 4, 049901.

Mashaah, M. T., Hakim, J., Chidzonga, M., Kangwende, R. A., Naik, Y., Federspiel, N., ... & Gomo, E. (2014). Strengthening research governance for sustainable research: experiences from three Zimbabwean universities. *Academic medicine: journal of the Association of American Medical Colleges*, 89(8 0), S69.

National Information Standards Organization. (2004). Understanding Metadata. https://www.lter.uaf.edu/metadata_files/UnderstandingMetadata.pdf

PHAIDRA. (o. D.). Phaidra. PHAIRDRA. <https://phaidra.univie.ac.at>

Re3data. (2022). Re3data. <https://www.re3data.org>

Tappeiner, G., Tappeiner, U., & Walde, J. (2007). Integrating disciplinary research into an interdisciplinary framework: A case study in sustainability research. *Environmental Modeling & Assessment*, 12(4), 253-256.

Tremouilhac, P., Lin, C. L., Huang, P. C., Huang, Y. C., Nguyen, A., Jung, N., ... & Bräse, S. (2020). The repository chemotion: infrastructure for sustainable research in chemistry. *Angewandte Chemie International Edition*, 59(50), 22771-22778.

Universität Wien. (o. D.)(a). *Forschung an der Universität Wien*. Abgerufen am 13. Dezember 2021, von <https://www.univie.ac.at/forschung/forschung-im-ueberblick/forschung-an-der-universitaet-wien/>

Universität Wien. (o. D.)(b). *U:Search*. u:search. Abgerufen am 13. Dezember 2021, von <https://usearch.univie.ac.at/>

Universität Wien. (o.D.)(c). Förderungen. <https://openaccess.univie.ac.at/foerderungen/>

Vukotich Jr, C. J., Cousins, J., & Stebbins, S. (2014). Building sustainable research engagements: Lessons learned from research with schools. *Journal of Research Practice*, 10(1).

Wirtschaftsuniversität Wien. (2016). Open Access Policy. WU Wien. <https://www.wu.ac.at/bibliothek/services/mitarbeiterinnen/forschungsunterstuetzung/open-access/open-access-an-der-wu/open-access-policy>

Wissenschaftsfonds zur Förderung der wissenschaftlichen Forschung. (o. D.)(a). Dashboard. <http://dashboard.fwf.ac.at/de/>

Wissenschaftsfonds zur Förderung der wissenschaftlichen Forschung. (o. D.)(b). Leitbild. <https://www.fwf.ac.at/de/ueber-den-fwf/leitbild>

Wissenschaftsfonds zur Förderung der wissenschaftlichen Forschung. (o. D.)(c). Open Access Policy. <https://www.fwf.ac.at/de/forschungsfoerderung/open-access-policy>

4. Appendix

4.1 Transcript Interview 1

Im Rahmen des Seminars beschäftigen wir uns mit nachhaltiger Forschung in dem Sinne, dass Daten häufig reproduziert und dementsprechend verschwendet werden. Das Wissen im Allgemeinen geht unter Umständen verloren. Meine Fragen an Sie: Was machen Sie an der WU? Inwiefern haben Sie Bezugspunkte zu diesem Thema? Was fällt Ihnen dazu ein?

A: Ich habe 2 Hauptaufgaben hier an der WU-Bibliothek: Fachreferat, also Informations- und Literaturbeschaffung im Bereich Mathematik, Statistik und Wirtschaftswissenschaften. Dieser Aufgabenbereich ist in den letzten Jahren allerdings immer geringer geworden... Der zweite Bereich ist das Thema Forschungsdatenmanagement. Damit beschäftigen wir uns seit 6 Jahren. Wir hatten ursprünglich mal die Aufgabe festzustellen: Welche Daten existieren hier überhaupt an der WU? Womit beschäftigen sie sich, welche Daten verwenden sie. WU lizenziert ja auch Daten kommerzieller Natur, die auch in der Forschung kostenpflichtig verwendet werden. Der Fokus ist natürlich im Wirtschaftswissenschaftlicher Bereich wie Banken, Börsen etc. Aber Wirtschaftsdaten ganz generell. Das Forschungsdatenmanagement ist eigentlich im Rahmen eines vom Wissenschaftsministerium österreichweite geförderten Projektes, das hat damals (unverst.) geheißen die Unis waren aufgefordert sich damit auseinanderzusetzen und die WU hat eine Strategie entwickelt. Wie gehen Forscher mit Daten um, wo gibt's Bedarfe seitens der Forscher, was brauchen die für Services für Speicherung, Forschungsanträge etc. Früher war es eine Beratung, wie stell ich einen Antrag beim FWF zum Beispiel... Um diese Strategie zu entwickeln muss ich wissen welche Daten die Forscher haben. Also haben wir uns 2016/17 die Daten anzuschauen, haben eine Publikationsanalyse gemacht und Interviews mit den Forschern aus diversen Departments zum Thema Datenmanagement, welche Datenkultur herrscht in den Departments. Wir haben heterogene Departments, natürlich wirtschaftswissenschaftliche Fächer aber auch Juristen, quantitative Forschung, qualitative Forschung wie Soziologie, Artefaktanalysen solche Dinge halt, auch Daten Produzieren, die

auch heikel sind. Das haben wir analysiert und sind darauf gekommen, dass wir eine Datenpolicy brauchen. Was soll die Uni anbieten, wozu verpflichtet sie sich und wozu sollen sich die Forscher verpflichten. Mittlerweile gibt's solche Datenmanagementpolicy an allen großen Unis in Wien. Hauptuni, WU, TU etc. ab 2016-2017 kam das.

F: So kürzlich?

A: Ja in den USA ist das seit 60er Jahren ein großes Thema, auch das Offenlegen der Daten. Hängt stark mit Genom und Medizin Forschung zusammen, die Druck machen. In den USA hat man früh begonnen in Datenrepositorien Daten für andere zur Verfügung zu stellen. Molekularbiologie, Genomforschung besonders. Bei uns in Europa ist das bissi später reingekommen. Das liegt auch damit zusammen, dass die EU sich entschlossen hat, einen gemeinsamen Datenmarkt haben zu wollen, wo auch öffentlich finanzierte Forschungsdaten eine große Rolle spielen. Wir haben dann diese Policy verabschiedet und seither beschäftigen wir uns mit dem Informationsaufbau, wir haben ein Schulungskonzept für den Herbst entwickelt. Außerdem gibt es das Forschungsservice beim Rektorat, die zum Thema informieren, ganz breit. Die IT, die die technische Seite beleuchtet. Das sind die drei Stakeholder quasi.

F: Wenn da die Policy ist, die 2016 erstellt wurde.

A: Die ist seit 2019 implementiert, im Anschluss auf diese Studie, das hat von 2017 bis 18 gedauert, das ging durch die Kommission, die Rechtsabteilung ob man das so schreiben kann, es ist ein politischer Prozess, es gibt Methoden, wie die Maßnahmen generiert werden und das wird geprüft. Es gibt weiche Policys, also Empfehlungscharakter. Zum Beispiel Forscher werden angehalten Open Access zu machen. Harte Policys sind dann obligatorisch.

F: Wenn man sich das konkret vorstellt, vor allem in Hinblick auf die Heterogenität... Wie kann man sich das konkret vorstellen? Ich bin ein Forscher an der WU, ich reiche ein Projekt ein und werde gefördert. Wie läuft das ab, wie kann ich mir das vorstellen?

A: Es gibt 2 Arten von Forschung: Auftragsforschung und oder man macht ein FWF-Projekt, da stellt man einen Antrag und das wird genehmigt. In der Regel muss der Antragsteller einen Datenmanagementplan ausfüllen, in dem er angibt, wie mit den Daten umgegangen wird, welche Art von Daten erstellt werden etc. Das ist auch bei WU-Projekten mittlerweile Standard. Man muss aber dazu sagen, die WU hat kein Datenrepositorium. Wenn jemand ein Paper schreibt, dann muss er die Daten extern bereitstellen. Da gibt es AUSSDA, Zenodo, das ist ein generisches Repositorium, da bekommt man einen digitalen Identifier für den Datensatz und dann hat man eine Referenz auf dem Datensatz. Es gibt keine durchgehende Einzelbetreuung bei einem Forscher, dass der konkret eine Ansprechperson hat, die er ansprechen kann. Wir haben im Forschungsservice eine Beratung im Rahmen von Forschungsförderungsanträgen oder auch Beratung im Bereich FWF, auch Angebote im Sinne von Auswahl von Förderprogrammen, Stipendien für PhD oder Habilitanden, die nach einer Förderung suchen. Das ist bei jedem Projekt sehr individuell die Beratung. Wir erfahren nicht von jedem Projekt. Manchmal brauchen Projekte lizenzierte Daten, die sie bei uns an der Bibliothek anfragen, aber diese Daten kann man nicht so einfach lizenzieren. Grundsätzlich gibt es aber auch viel Forschung außerhalb der geförderten Bereiche. Als wir der Publikation gemacht haben, waren 80% nicht gefördert. Der Grund warum wir die Studie gemacht haben, war weil wir keine Ahnung haben was für Daten da erhoben werden. Wir haben erfahren, dass Juristen eigentlich keine Forschungsdaten produzieren, zumindest nicht standardmäßig. Die arbeiten nicht empirisch, daher waren sie nicht der Mittelpunkt unserer Betrachtung. Wir haben ein Repositorium, aber nicht für Forschungsdaten sondern für Veröffentlichungen, das heißt ePub. Dort kann man auch unfertige Arbeiten ablagern. Das gibt es seit 2008 oder so. Der Betrieb eines Repositoriums ist sehr aufwendig, dafür braucht man viel Manpower für die Implementierung, auch wenn es Open-Source-Möglichkeiten gibt, die vom Ankauf her kein Problem wären. Es gibt in Ö ein Fair-Data Projekt, an dem Universitäten arbeiten auf Basis von dem Invenio von CERN einen Prototypen entwickeln, der aber für alle Forschungsdaten im Grunde geeignet wäre. Auf Rektoratsebene haben wir uns meines Wissens nicht damit auseinandergesetzt. Forschern

wird immer empfohlen, ihre Daten in fachspezifischen Repositorien abzulegen, da gibt es viele verschiedenen. Die Forscher legen die Daten eher dorthin als auf allgemeinen Repositorien. Es gibt einen Bedarf an Ablagemöglichkeiten für kleine Datensätze, die man nicht in einem fachspezifischen Repository ablegen muss, es gibt aber auch sehr große Datenmengen, die nicht ohne weiteres aufgrund der Datenstruktur gehostet werden können. Ich glaube dass die Frage, ob die Uni ein eigenes Repository braucht, nicht so leicht beantwortet werden kann. Die Uni Wien hat PHAIDRA, weils auch größer ist, es ist eine Volluni. Wir haben schon die Möglichkeit, Daten bei uns zwischenspeichern, da kann man mit der IT sprechen, das kostet natürlich. Aber für Langzeitspeicherung haben wir kein eigenes Repository. Ob es für die Zukunft so bleibt kann ich nicht sagen. Derzeit steht es nicht auf der Agenda was einerseits auch damit zusammenhängt, dass wir ein neues Forschungsinformationssystem bekommen, das gerade implementiert wird, da könnte man ein Datenrepository dranhängen. Das gehört allerdings nicht zur politischen Positionierung des Rektorats derzeit. Eine zweite Möglichkeit ist es, wenn man kein eigenes möchte, dann kann man sich an einem Repository bedienen, was es bereits gibt. AUSSDA ist da so ein Fall. Da hätten die Forscher auch ihr Zielpublikum und da besteht die Möglichkeit, Mitglied zu sein, aber das kostet natürlich wieder. Aber es könnte sein, dass man da jemanden sitzen hat, aus den eigenen Reihen, der sich um die Forschenden kümmert.

F: Die Umsetzung der Datenmanagement ist auf Forscherseite?

A: Genau, die können sich an uns oder das Forscherservice melden, sie können aber auch, das ist das was wir in den österreichweiten Projekten versucht haben, jede Uni muss nicht das selber machen und neu erfinden, sondern dass man.. Durch Covid gab es viele Onlineveranstaltungen auch im Rahmen der Projekte und da hat sich gezeigt dass es gut ist wenn man diese Beratung hat, das muss nicht jede Uni haben, das kann auch ein consoziales Service sein, dann kann der Forscher das in Anspruch nehmen. Natürlich ist die Nähe zur Universität, wenn man sich an jemanden wenden kann, der mit Eigentümlichkeiten der

Universität vertraut ist, von Vorteil. Ich hatte gestern so ein Gespräch zu dem Thema, da geht es ja nicht nur um den Inhalt, sondern auch um Richtlinien und jede Universität hat eigene Richtlinien, wer darf zugreifen. Unterschiedliche rechtliche Rahmenbedingungen auch in internationalen Projekten. Da ist es gut, wenn man das auf der eigenen Uni hat. Häufig fehlt auch Personal. Eine Abteilung zu haben, die das alles managet geht eigentlich nicht so wirklich. Wir sind zwar zentraler Ansprechpartner, aber eine verbindliche Aussage kann nur die Rechtsabteilung geben.

F: ePub haben sie angesprochen, können Sie das genauer erklären? Ist das nur ein Speicherungssystem oder auch dazu geeignet relevante Literatur zu suchen?

A: Ja es gibt eine Suchfunktion aber Sie finden das auch über diverse Suchmaschinen. Der Bibliothekskatalog sucht ePub auch ab, die Metadaten werden mitindexiert. Teilweise sind die Workingpaper auch in internationalen Portalen suchbar. Nicht nur zum Teil, aber in der Ökonomie gibt es internationale Portale, die Workingpaper-Series hosten, da werden die Daten hier und da gefunden. Im Prinzip ist alles in ePub auch auffindbar im Katalog.

F: Open-Access ist ein großes Thema, wie sieht es da aus, Entwicklungen hin zu, oder sogar weg von.

A: Muss man schauen. Also Daten, wenn man kommerzielle Daten anschaut von Reuters, Bloomberg etc., die sind immer zahlungspflichtig, auch für die Unis. Die Unis sind natürlich nicht die primären Zielgruppen, aber es wird viel damit gearbeitet. Aber es gibt auch Mikrozensus etc. die öffentlich zugänglich sind bei Statistik Austria etc.

F: Gibt es in der Policy den Anspruch, dass es open-access sein muss?

A: Es gibt eine Open-Access-Policy, da wird empfohlen möglichst offen zu publizieren, auch die Daten. In den Top-Zeitschriften ist es so, dass auch die Journals selbst die Data-Policy haben, die sehr stark das Zugänglichmachen der Forschungsdaten empfehlen oder

manchmal auch fordern. Das hat verschiedene Hintergründe, besonders in der quantitativen Forschung geht es auch um die Reproduzierbarkeit, um Fälschungen zu verhindern. Anderer Grund ist die Erhebung von Primärdaten, die ist sehr kostspielig. Wenn ich zu dem Thema aber einen Datensatz habe, dann ist es leicht, diese Daten wiederzuverwenden mit anderer Forschungsfrage. Es gibt aber auch Daten, die so speziell sind, dass man sie nicht wiederverwenden kann. Da macht es keinen Sinn, sie öffentlich zu machen.

F: Was spricht aber dagegen?

A: Da spricht eigentlich nichts dagegen, es gibt halt keinen Nutzen. Es sei denn man sagt als Journal, dass man die Daten offenlegen soll, damit man die Forschungsergebnisse reproduzieren kann.

F: Es gibt an der WU die Empfehlung für Open-Access, es ist aber nicht obligatorisch?

A: Nein ist es nicht, das hängt stark von den Disziplinen ab. Es kommt auch drauf an, was die Verlage sagen, in welchen Publiziert wird. Es macht keinen Sinn, wenn die Uni sagt ihr müsst Publizieren, das ist ja mit viel Aufwand verbunden und großes Problem ist das Wertesystem. Wenn ich für den Aufwand um die Daten zu publizieren, keinen Impact hab. Wenn ich sag in meinem Curriculum bekomme ich keinen Punkt, keinen Bonus, dass ich die Datenpublikation habe, dann wird der Forscher den Aufwand da minimieren, da er eh gezwungen ist, viel zu Publizieren. Das war ein wesentliches Resultat unserer Studie, dass die einzelnen Fachbereiche sehr unterschiedliche Zugänge zur Offenheit haben. In der Mathematik ist es gang und Gäbe, dass man Daten teilt. In anderen Bereichen ist es nicht so üblich. Oft ist der Datensatz für den Forscher ein Alleinstellungsmerkmal. Wenn er mit den Daten noch weiter Publizieren kann, dann wird er die Daten nicht so gern hergeben, denn es herrscht großer Wettbewerb in der Forschung. Man kann es leider nicht so einfach vereinheitlichen, weil die Forschungskulturen in den einzelnen Disziplinen sehr unterschiedlich ist.

F: Liegt es am Status Quo oder an dem Level an Wettbewerb.

A: Wettbewerb ist dort, wo ich einen Datensatz hab. Wenn ich weiß ich hab 5-6 Jahre an dem Datensatz gearbeitet und ich kann 5-6 Paper damit veröffentlichen, dann wird ich das nicht veröffentlichen. Gut im Nachhinein vielleicht fürs Journal aus Transparenzgründen oder weil es verlangt wird, aber nicht für alle. In der Wissenschaft lebst du davon, dass dein Paper in einem Guten Journal ist, für deine Habilitation etc. Wir haben ein Wertesystem, das auf diesen Zeitschriften aufbaut. Deine Karriere hängt davon ab.

F: Fachspezifischen Repositorien: Sind die noch zeitgemäß bei heutigen komplexen Fragestellungen?

A: Ja, die haben auch sehr lange Tradition, nehmen wir her es gibt diese Genom-Datenbank, der Vorteil ist, dass die Daten die da eingespielt werden sehr homogen sind, das erhöht den Wert des Datenbestandes. Wenn ich ein Archiv habe mit verschiedenen Daten wie Bild und Software, ist es heterogen, das hat man in generischen Repositorien wie Zenodo. Das hat den Vorteil, man kann sehr schnell seine Daten reinstellen und bekommt schnell einen DOI. Bei einfachen Daten reicht das. Wenn die Datenspeicherung betreut wird, wie bei AUSSDA, geht es nicht so schnell und einfach. Kommen wir zur Datensuche: Da profitiere ich von fachspezifischen Repositorien. Die Datenbeschreibungen basieren auf Standards, die je nach Fach unterschiedlich sind. Die Daten werden mit Metadaten beschrieben. Die stimmen nicht miteinander überein. Das macht es bisschen schwierig, wenn man über alle Fächer hinweg suchen möchte.

F: Es gibt 3 Repositorien: Fachspezifische Repositorien, mit fachspezifischen Datenstrukturen. Dann gibt es generische Repositorien wie Zenodo und es gibt institutionelle Repositorien, die nur für angehörigen einer Institution gedacht sind. Letztere werden wahrscheinlich am ehesten verschwinden. Die haben andere Zwecke und ergänzen sich. Institutionelle ermöglichen es, dass man schnell etwas hochladen kann und man hat

eine gewisse Datenhoheit, da spielt was politisches natürlich eine Rolle. Wenn die Uni etwas finanziert, dann will sie natürlich mitreden. Dann will man es nicht auf Zenodo haben.

F: Die Daten suchen: Was gibt es da an der WU?

A: Forschungsdaten sucht man nicht im Unibibliothekskatalog. Wenn man Daten sucht gibt es beispielsweise Base, das ist ein deutsches Produkt, dort kann man nach Forschungsdaten suchen. Ob da alles drinnen ist, das ist nicht nachvollziehbar. Ein Forscher in einem Fachgebiet weiß wen er fragen muss, um bestimmte Daten zu finden, oder der weiß in welchen Repositorien er sucht. Wenn das aber unklar ist, dann ist es von Zufall geprägt. Zum Beispiel hab ich ein Interview gehört: sie haben ein Hotelsujet jedes Jahr fotografiert und archiviert über 30 Jahre oder so. Dann hat ein Gletscherforscher dieses Foto gesehen und im Hintergrund einen Berg gesehen. Auf jedem Foto war der Zustand des Gletschers anders. Der hat diese Bilder genommen und analysiert. Das war Zufall. Bei der Suche ist das Problem, dass man nicht weiß ob es den Datensatz bereits gibt, ein zweites Problem ist die Verknüpfung von Daten, die frei zugänglich sind in Kombination mit Daten, die ich lizenzieren muss. Dann habe ich Daten die aus zwei Teilen bestehen und dann habe ich das Dilemma, wenn ich es Publizieren will, dass ich das nicht kann. Aber in der Suche kann man immer etwas verbessern. Ich glaube, dass forscher wissen, wo sie suchen müssen, weil sie im regen Austausch untereinander stehen. Die Suche nach einem Datensatz ist schwierig, wenn man nicht weiß wie sie beschrieben sind. Man braucht jemanden, der weiß, wie die Daten beschrieben sind. Die Suchstrategie ist individuell, viele googlen einfach, aber einen Datensatz findet man so eher nicht. Aber es gibt einige gute Indexierungen, z.B Base. Aber auch r3data ist sehr gut, da sucht man drinnen, wenn man nach einem Archiv sucht, in welchem man die Daten veröffentlichen möchte. Dort kann man schauen, ob es Repositorien gibt, die zu meinem Thema sind und kann dort schauen, ob ich was finde. Man braucht eine gute Indexierung, damit man findet was es gibt. Am besten wäre ein Google für Daten. Aber es setzt voraus, dass die Daten gut beschrieben sind, das schreiben in der

Regel die Forscher selbst. Je besser die Metadatenbeschreibung, desto besser zu finden und zu nutzen mit andern. Mit dem steht und fällt das Ganze.

F: Wenn jemand an der WU was publiziert, unterstützt man ihn auch bei der Metadatenbeschreibung?

A: Wenn sich jemand bei uns meldet dann ja. Es gibt so Standards, die generisch sind, aber es gibt dann natürlich fachspezifische Daten. Klassischer Metadaten-standard der Fachunabhängig ist ist der Dublin-Core. Der lässt sich immer Verwenden, ist aber allgemein. Die, die viel mit Daten arbeiten, die finden sich dann natürlich leichter zurecht. Wenn sie zu einem Fachrepositorium gehen, dann bekommen sie vom Repositorium Hilfe. Das ist dezentralisiert. Bei AUSSDA oder so hat man immer eine Person, die einen unterstützt.

4.2 Transcript interviewee 2

F: Ich würde Sie bitten sich kurz vorzustellen und mit mitzuteilen, was Sie machen.

A: Mein Name ist NAME NACHNAME ich bin seit JAHR an der UB Wien tätig und habe Phaidra mit aufgebaut, das Repositorium der Uni wien. Basierend darauf sind andere Aufgaben wie Forschungsdatenmanagement, Datenmanagementplanberatung und so weiter. Ich bin Germanistin und auf einem Germanistischen Institut tätig als Lektorin und bin in das Thema Forschungsstätten dann hereingeschlittert.

F: Wir wollen nachhaltige Maßnahmen am Institut umsetzen, wir haben bisher Literatur recherchiert und ich habe herausgefunden, dass ein Teil der Forschung, der nicht nachhaltig ist dass sehr viel reproduziert wird, was vermieden werden kann durch das Teilen von Daten. Wie werden Daten an der Uni Wien gespeichert, wie ist der Prozess?

A: Spannendes Thema, die Nachhaltigkeit der Daten wird in den letzten Jahren immer wichtiger bei allen Universitäten aber auch bei den Fördergebern. Seit 2019 fordert der FWF einen Datenmanagementplan bei welchem man vorlegen muss, wie man mit den Daten, mit denen man arbeitet, umgeht. Früher war es so, dass wenn man ein Projekt

abgeschlossen hat, hat man eine Publikation gemacht und kein Mensch hat nach den Daten gefragt, es war egal was mit den Daten ist. Jetzt muss man in jeder Fachrichtung angeben, was mit den Daten die man erzeugt und verwendet gemacht wird. Was auch wichtig ist, ist welche Daten man verwendet. Daten sollten ja weiterverwendet werden. Da liegt der Fokus sehr stark drauf aktuell, aus mehreren Gründen. Einerseits will man Geld sparen, andererseits will man es transparenter und nachvollziehbar gestalten beispielsweise mit Peer-Reviews. Früher hat man eine Publikation gehabt und musste ihr vertrauen, die Art und Weise, wie die Autoren auf das Ergebnis kommen war nicht klar, das wird immer wichtiger, diese Transparenz.

F: Die haben den Datenmanagementplan angesprochen, das hört sich so an als müssten sich die Forscher selbst darum kümmern, habe ich das richtig verstanden?

A: Ja grundsätzlich ist es so, dass ForscherInnen aufgefordert werden so einen Managementplan abzugeben, aber viele haben sich damit noch gar nicht auseinandergesetzt. Es gibt viele Fragen die vor 10 Jahren noch nie ein Thema waren und mit denen kommen ForscherInnen auf uns zu und benötigen Unterstützung. Wir sind ein Team und wir versuchen dort zu helfen wo Fragen entstehen.

F: Wie funktioniert die Datenspeicherung an PHAIDRA.

A: Es gibt auch AUSSDA, das ist das Repositorium für Sozialwissenschaften, alles andere kommt zu PHAIDRA, es gibt auch andere aber die sind die der Uni wien.

F: Warum gibt es nicht ein zentrales

A: Gute Frage, warum gibt es kein zentrales oder auch österreichisches Repositorium? In anderen Ländern gibt es nationale Repositorien... Es hat sich so nicht ergeben, es gab mal Gespräche zu einem nationalen Repositorium, das wurde aber abgelehnt, also hat jede Uni versucht ein eigenes aufzubauen bzw. mitzunutzen. Bei PHAIDRA haben wir 20 Partner, die auch PHAIDRA verwenden, wie FH St Pölten, Landesbibliothek Vorarlberg, FWF Kunstuni

Graz und Linz sind Kooperationspartner. Es gibt aber viele andere Systeme die verwendet werden, die unterschiedliche Funktionen haben. Bei uns soll PHAIDRA für alles sein. Jedes Objekt, das digital ist egal ob Fotos, Dokumenten oder Tonaufnahmen. Es gibt spezialisierte Repositorien aber auch, manche beginnen mit Dokumenten und fügen später andere Daten noch als Möglichkeit hinzu. Zur Frage wie das funktioniert ist unterschiedlich. In AUSSDA geben sie die Daten ab, die werden curatiert, die Metadaten werden ergänzt und dann werden sie hochgeladen und geprüft und am Ende freigegeben. Bei PHAIDRA kann jede Person, auch Studierende, was hochladen. Es gibt Nutzungsbedingungen aber wir werden nicht kontrollieren, wie gut die Qualität der Daten sind, das können wir uns auch nicht leisten. Bei uns ist alles würdig zu archivieren. Wir unterstützen die Forscher natürlich aber wir laden es nicht für sie hoch und sie müssen auch die Metadaten selbst eintragen.

F: Bei Fragen bekommt man aber trotzdem Untertützung?

A: Natürlich. Und wenn es viele Daten sind bieten wir einen alternativen Upload mit dem ZID an. Wir arbeiten eng mit ihm zusammen.

F: Wenn man einen Account hat kann man einfach daten hochladen. Braucht man einen account um die Daten runterzuladen?

A: Die Person die hochlädt kann einstellen wer die Daten sehen darf. Es gibt die Möglichkeit, Daten zu sperren. Sie haben ein Forschungsprojekt für 3 Jahre und Sie können einstellen, dass für diese 3 Jahre keiner sie sehen kann, weil Sie arbeiten mit diesen Daten also soll niemand was damit machen. Oder sie haben es auf 20 Jahre gesperrt und dann sind sie frei verfügbar. Andererseits können Sie auch eine Lizenz draufgeben zB nicht für kommerzielle Zwecke verwendet werden. Oder sie sind Open Access unterwegs. Jeder kann dann auch runterladen, ohne Zugang.

F: Gibt es eine Barriere abgesehen von Lizenzen, z.B. Regeln dass man es nur machen kann wenn man an der Uni wien ist?

A: Nein bei uns ist das nicht so. Wir können sagen die Daten sind gesperrt aber wenn sie die daten für wissenschaftliche zwecke benutzen dann melden Sie sich bei uns. Das ist nützlich bei Daten, die rechtlich nicht für Open Access geeignet sind. Bei AUSSDA gibt es eigene Lizenzen für die wissenschaftliche Forschung, wo daten nur für diese Zwecke genutzt werden können.

F: Thema ist auch Open Access, es sollte möglichst Open Access sein, haben Sie eine einschätzung, wie viele Ihrer Daten open access sind?

A: Hab ich lang nicht geschaut, man kann bei PHEIDRA einfach im Suchfeld nachschauen, das ist transparent aufgelistet, auch die Lizenzen.

F: Wie funktioniert das suchen intern? Normalerweise kann man nach Suchbegriffen Filter nutzen, um relevante Ergebnisse zu finden. Dabei ist es immer wichtig, wie die Daten kategorisiert sind etc. Wie läuft das bei PHEIDRA ab? Wie werden die Metadaten kategorisiert, was ist das Metadatenschema?

A: Die Daten werden beschrieben und zugleich so, dass sie maschinen auch lesen können. Die Kategorisierung erfolgt durch die Datenbeschreibung. Wir haben verschiedene Klassifikationen, Beschreibmöglichkeiten, nach Begriffen nach Lizenzen, es gibt verschiedene Möglichkeiten.

F: Und wer kategorisiert das?

A: Die Forscher selbst

F: Problem ist ja auch, dass es für Konzepte manchmal, besonders International, unterschiedliche Namen gibt. Da muss man also hoffen, dass die Forschenden es möglichst treffsicher angeben.

A: Genau, die Forscher bekommen natürlich Hilfen, wir greifen aber nicht ein, das können wir auch nicht. Es ist so, wir konzipieren die Oberfläche von PHEIDRA gerade neu. Dann wird

es eine Anbindung an die GNP geben an die (unverst.) Das ist ein bestimmtes, überprüfbares Vokabular. Bis jetzt war es nur marginal der Fall, das funktioniert aber ganz gut weil wenn wir an Projekte denken, die Projektleiter geben vor wie es genannt wird nach wissenschaftlichen Kriterien und das wird bei Pheidra auch angezeigt.

F: Wenn wir davon ausgehen, dass ein Datensatz häufiger genutzt wird, gibt es eine Möglichkeit das Datenfile auszuwählen und zu den Daten verlinkt zu werden?

A: Das gibt es nicht, das kommt vielleicht in Zukunft aber dass man sieht wo der Datensatz zitiert wurde, aber das gibt es derzeit nicht, ist schwierig auch Datenschutzrechtlich.

F: Abseits von dem was Sie angesprochen haben, was sind weitere Ideen um die Nutzung von Daten weiter zu befördern?

A: Baustellen gibt es genug, wir denken an sehr viele neue Ideen bei Pheidra. Der Re-Use von Daten wird bei uns gerade neu überlegt. Bis jetzt waren wir auf der Sichtweise daten speichern und archivieren. Das war mal der erste schritt, was mit den Daten passiert liegt an den Forschenden. Jetzt in einem Zweiten Schritt denken wir an den Reuse und schauen uns an was die Forscher brauchen. Das ist der nächste Schritt und dann überlegen wir uns ob wir neue Dinge entwickeln oder ankaufen. Vielleicht zusätzliche Features, es ist ja ein Open-Source Produkt aber es kann gut sein, dass wir was dazukaufen um es zum Beispiel mit vorhandenen produkten zu verknüpfen, aber so weit sind wir noch nicht,

F: Worin liegt das größte Problem daran, die Daten zu Reusen.

A: Ich glaube rechtliche Fragen sind häufig das problem, besonders bei Geisteswissenschaften. Die dürfen oft nicht archiviert oder weitergegeben werden, da sind wir weit weg von OA. Ethische Fragen sind auch oft das Problem, darf ich das noch verwenden heutzutage oder nicht. Früher war es für eine Handvoll Leute heute ist es für die ganze Welt. Das sind Dinge die jetzt erst richtig diskutiert werden, wir stehen bei der Datenarchivierung und Datenweaternutzung ja eigentlich noch ganz am Anfang.

F: Wann hat diese Entwicklung begonnen?

A: Eigentlich haben wir sehr früh begonnen, wir beschäftigen uns mit der Archivierung seit 2007. Aus dem Grund, dass man nicht möchte das die Daten verschwinden, hat man PHEIDRA ins Leben gerufen. Es war der Fokus auf einer langfristigen, sicheren Datenarchivierung. Das ist auch nicht trivial, man muss sicher sein, dass man ein PDF-Dokument auch in 50 Jahren noch öffnen kann. Früher war es so, dass man ein Buch auch in 200 Jahren noch lesen kann, solange kein Feuer ausbricht. Digital kann es passieren, dass etwas nicht mehr funktioniert. Wenn man an Nachlässe denkt, ein Autor muss 70 Jahre verstorben sein bevor das Dokument gemeinfrei ist. Bei Interviews sind die rechtlichen Fragen sehr wichtig. Ob ein Interview aufgenommen werden darf, transkribiert werden darf und veröffentlicht wird benötigt die Zustimmung des Interviewpartners. Diese kann allerdings auch wieder widerrufen werden, darum muss man sich auch kümmern. Bei Bildern muss man beachten wer das Bild aufgenommen hat, das weiß man bei älteren Bildern oft nicht. Da wurden Bilder archiviert und nach 20 Jahren gefunden, aber dann darf man die Bilder nicht nutzen, weil man nicht weiß ob das der Fotograf möchte. Heute wird das alles berücksichtigt, früher aber nicht. Die digitale Welt ist ja nicht so alt. Die rechtlichen Fragen sind teilweise sehr detailreich, da kommen unterschiedlichste Fragen auf uns zu, wir sind ja keine Juristen. Wir geben das dann weiter aber da gibt es eine Hürde.

F: Wenn die Regulation zugenommen hat, wie wird ihrer Meinung nach die zukünftige Entwicklung sein?

A: Es gibt manche Menschen ihren Lebensunterhalt damit verdienen. Vor allem in Deutschland gibt es Rechtsanwälte, die nach solchen Themen suchen, die nicht rechtens sind und senden dann eine Klage. Die werfen uns dann vor, dass wir Dokumente rechtswidrig archivieren und verlangen hohe Beträge. Das ist uns schon passiert, wir mussten zwar nie etwas zahlen, aber es wurde uns angedroht. Das zeigt, dass da auch ein

großer Markt ist. Ich kann mir schon vorstellen, dass man strenger wird und klar reglementiert wird.

F: Bedeutet das auch, dass das Teilen von Daten schwieriger wird?

A: Man muss bewusster damit umgehen, also beispielsweise hat man früher irgendwelche Bilder in eine Powerpoint Präsentation gegeben und diese dann geteilt. Das geht heute nicht mehr, da muss man bewusster werden. Bei Vorträgen werden die Bilder oft live gezeigt, in der Archivierung werden die Bilder dann aber rausgenommen. Da sieht man dass das Bewusstsein steigt, auch bei Hochschulschriften wie Dissertationen. Ich glaube dass das ein Prozess ist, der nicht abnehmen wird. Es gibt Bestrebungen, dass sich das Recht an die digitale Welt anpasst, es soll mehr in geordneten Bahnen laufen.

F: Wie finden Wissenschaftler heraus, dass es PHEIDRA gibt, wie wird PR betrieben?

A: Das ist eine wichtige Frage auch für uns. Wir werden zum Beispiel auch auf Google gefunden, also die Objekte. So findet man auch unsere Objekte. Unsere offenen Dokumente sind auch auf u:search zu finden. Es gibt mehrere Möglichkeiten darauf zuzugreifen. Aber bekannt zu werden ist eine Sisyphus-Arbeit. Es gibt die Forschungsdaten-Policy an der Hauptuni für die Forschenden damit sie wissen was sie tun müssen, wenn sie an der Uni Wien sind und Forschungsdaten produzieren, da sind wir grad dabei das bekannter zu machen. Es ist aber schwierig, weil die uni sehr hierarchisch ist, mit verschiedenen Ebenen und es ist schwer die Gesamtheit zu erreichen. Wir versuchen das mit Institutskonferenzen und damit, dass wir aktiv auf die Forschenden zugehen. Aber es ist nicht einfach, das Personal fluktuiert auch stark. Es wird aber immer bekannter und die Zugriffszahlen steigen an, durch Mundpropaganda wird es auch bekannter gemacht.

F: Was ist ihrer Meinung nach die beste Möglichkeit zu verhindern, dass Daten unbeabsichtigt dupliziert werden?

A: Ich glaube, dass die Forschenden lernen müssen herauszufinden, was es schon gibt. Das ist in einigen Disziplinen schon sehr gut ausgeprägt, bei anderen Fächern allerdings nicht. Man muss wissen wo gibt es die Daten und wie kann ich die verwenden. Bei Geisteswissenschaften ist das nicht sehr bekannt, das kann ich von Erfahrung sagen. Das ist ein Lernprozess, der noch im Werden ist. Es gibt auch in den Geisteswissenschaften natürlich Personen, die da sehr affin sind, aber auch das Gegenteil. Das wird sicher noch dauern. Früher war das Denken auch so: Das sind meine Daten, mit denen darf nur ich arbeiten und ich arbeite nicht mit anderen Daten arbeiten, weil auf die kann ich mich nicht verlassen.

F: Es gibt ja unzählige Datenbanken, sowohl national als auch international. Gibt es Vorteile an der Zahl der Datenbanken?

A: Ich glaube das sind politische Gründe, es hat keinen logischen Grund. Weil wenn man ein großes Repositorium hat, was zentral verwaltet wird wie Zenodo von Cern, wenn wir sowas zentral hätten, dann wäre es einerseits ganz gut, auf der anderen Seite gibt es immer Sonderwünsche. Je nach Disziplin gibt es unterschiedliche Anforderungen an das Repositorium. Damit kann man eine zentrale Datenbank vergessen. Es findet auch kein Austausch zwischen Unis statt. Es gibt ja das Projekt „European Open Science Cloud“ da soll ja die Metasuche passieren, die auf alle Repositorien zugreifen kann. Ein bisschen wie Euroklana macht, wo es unterschiedliche Daten aus unterschiedlichen Repositorien gibt und die dort angezeigt werden und auf die man zugreifen kann. Das sehen wir in der Zukunft, dass man Repositorien darauf vorbereitet, dass es Schnittstellen gibt auf die man auch extern zugreifen kann.