YULIA KARIMOVA

RESEARCH DATA DESCRIPTION IN MULTIPLE DOMAINS: SUPPORTING RESEARCHERS WITH DATA MANAGEMENT PLANS

SUPERVISOR: MARIA CRISTINA DE CARVALHO ALVES RIBEIRO (PH.D.)
CO-SUPERVISOR: GABRIEL DE SOUSA TORCATO DAVID (PH.D.)

DOCTORAL PROGRAM IN DIGITAL MEDIA

Faculty of Engineering of University of Porto
http://www.fe.up.pt
2023

Yulia Karimova: *Research data description in multiple domains: supporting researchers with data management plans,* Doctoral Program in Digital Media, 2023.

ORCID:
https://orcid.org/0000-0002-1015-6709

E-MAIL:
ylaleo@gmail.com

ABSTRACT

The growing amount of research data and the recognition of their value in the scientific community has led to the development of policies, infrastructures, tools and services. Researchers are encouraged to open their data and make them FAIR (Findable, Accessible, Interoperable and Reusable) to ensure the verifiability of results and transparency of research, while increasing their value. In this context, Research Data Management (RDM) becomes a vital component of high-quality research and a regular activity for researchers. The Data Management Plan (DMP), provision of quality metadata, repository selection, and data publication are some of the activities that researchers need to undertake when following the Open Science policies. However, these activities are time-consuming and not obvious to researchers, requiring specific knowledge, effort and experience. To support researchers, institutions are looking for solutions and developing workflows, tools, systems and services that aim to meet RDM requirements of both researchers and funding agencies.

Implementing any information system, tool or service at a research institution requires their development, testing and evaluation, as well as adaptation to existing regulations. In this work, we present the development of a DMP support system that meets RDM requirements and researcher needs. We also propose its implementation in the institutional infrastructure and workflow. We explore various aspects related to the DMP support system development, namely RDM and DMP requirements and legislation at the international and national levels, institutional project management and RDM workflows, existing initiatives, recommendations and tools.

We introduce a collaborative DMP-building method based on collaboration between researchers and a data steward. We describe its application to case studies from different scientific domains, analyze the results and propose its systematic evaluation. We also propose a path that simplifies the DMP creation process, to make plans more standard, high-quality, and detailed, by implementing controlled vocabularies in the DMP support system. To this end, we apply the collaborative method on projects from the same scientific domain, identifying specific aspects for this domain and proposing terms that can be used with controlled vocabularies and possibly contributing to the design of Domain Data Protocols.

The proposed DMP support system also considers the existing mechanisms of automation and interoperability, namely the machine-actionable DMP (maDMP) standard. We analyze the maDMP standard and existing tools created according to this standard and propose a DMP structure and a DMP workflow that can help to make the DMP machine-actionable. Finally, we suggest the implementation of the DMP support system in the institutional RDM workflow, and the interface to support researchers in DMP creation and RDM activities in general. Moreover, this work concludes that the DMP support system can be implemented in any institution, and used for projects in different scientific domains. The results constitute a guide for institutions that are still at the beginning of the development of the RDM infrastructure.

The feedback from researchers shows that they need support in DMP creation and monitoring, not only at the planning stage of the project but also during its course and sometimes even after. They are very positive about the implementation of the proposed system, highlighting its importance and necessity. Along with researchers, institutions also benefit from the implementation of such a system by providing support to researchers in RDM issues, reducing the time to create plans, improving the quality of DMPs and making projects comply with the RDM and funders' requirements, thereby contributing to the prestige of the institution.

RESUMO

A crescente quantidade de dados de investigação e o reconhecimento do seu valor na comunidade científica levou ao desenvolvimento de políticas e infra-estruturas. Os investigadores são incentivados a abrir os seus dados para assegurar a verificabilidade dos resultados e a transparência da investigação. Neste contexto, a Gestão de Dados de Investigação (RDM) torna-se uma componente vital da investigação de alta qualidade e uma actividade regular para os investigadores. O Plano de Gestão de Dados (DMP), o fornecimento de metadados de qualidade e a publicação de dados são algumas das actividades requeridas para seguir as políticas de Ciência Aberta. Contudo, estas actividades são demoradas e não óbvias para os investigadores, exigindo conhecimentos específicos, esforço e experiência. Para apoiar os investigadores, as instituições estão à procura de soluções e a desenvolver ferramentas, sistemas e serviços que visam satisfazer os requisitos do RDM.

A implementação de qualquer sistema, ferramenta ou serviço de informação numa instituição de investigação requer o seu desenvolvimento, teste e avaliação, bem como a adaptação aos regulamentos existentes. Neste trabalho, apresentamos o desenvolvimento de um sistema de apoio para DMP que satisfaz os requisitos do RDM e as necessidades dos investigadores. Propomos também a sua implementação na infra-estrutura institucional. Exploramos vários aspectos relacionados com o desenvolvimento do sistema, nomeadamente requisitos e legislação do RDM e DMP a nível internacional e nacional, gestão de projetos institucionais e fluxos de trabalho do RDM, iniciativas, recomendações e ferramentas existentes.

Introduzimos um método de construção de DMP baseado na colaboração entre investigadores e um *data steward*. Descrevemos a sua aplicação a casos de estudo de diferentes domínios científicos, analisamos os resultados e propomos a avaliação sistemática. Propomos também a criação do DMP através da implementação de vocabulários controlados no sistema de apoio para DMP, para fazer planos mais padronizados, de alta qualidade e detalhados. Para tal, aplicamos o método de colaboração em projetos do mesmo domínio científico, identificamos e propomos termos específicos que possam ser utilizados nos vocabulários controlados e contribuir para o desenvolvimento de *Domain Data Protocols*.

O sistema de apoio para DMP proposto considera os mecanismos existentes de automatização e interoperabilidade, nomeadamente a norma *machine-actionable* DMP. Analisamos a norma e as ferramentas criadas de acordo com a mesma e propomos uma estrutura de DMP que possa ajudar a tornar os DMPs accionáveis por máquina. Sugerimos a implementação do sistema de apoio para DMP no fluxo de trabalho institucional, e a interface para apoiar os investigadores na criação de DMP e nas actividades de RDM em geral. Além disso, este trabalho conclui que o sistema de apoio para DMP pode ser implementado em qualquer instituição e utilizado para projetos em diferentes domínios científicos. Os resultados constituem um guia para as instituições que ainda se encontram no início do desenvolvimento da infra-estrutura do RDM.

O feedback dos investigadores mostra que necessitam de apoio na criação e monitorização do DMP, não só na fase de planeamento do projeto, mas também no decurso deste e por vezes mesmo depois. Eles dão-nos reforço quanto à implementação do sistema proposto, sublinhando a sua importância e necessidade. Juntamente com os investigadores, as instituições também beneficiam da implementação de tal sistema, prestando apoio aos investigadores em questões de DMP, reduzindo o tempo para criar planos, melhorando a qualidade dos DMPs e fazendo com que os projetos cumpram os requisitos do RDM e dos financiadores, contribuindo assim para o prestígio da instituição.

ACKNOWLEDGEMENTS

In carrying out this PhD work, I had the direct or indirect support of multiple people, institutions, God, my guardian angels and other mysterious forces to which I am deeply grateful. At the risk of unfairly not mentioning any of the contributions, I want to express my thanks:

To the supervisor of this thesis, **Prof. Maria Cristina de Carvalho Alves Ribeiro**, for the guidance provided, for encouragement, availability, patience, confidence, wisdom and disposition to help. Here I express my immense gratitude to you.

To the co-supervisor **Prof. Gabriel de Sousa Torcato David**, for his availability, encouragement, knowledge and also for support in preparing this work.

To my mom **Nina Petrovna Karimova (Shumilova)** and my dad **Rinat Asgatovich Karimov** for everything, for all the moments in my life, for everything that I had, have and will have, for their support, at any time, at any moment, whether it snows or rains, for that their presence whenever I need them, and for the fact that they are the best parents ever!

To all friends and colleagues who, directly or indirectly, contributed, helped, and motivated me in the preparation of this study, for their patience, attention and strength they provided at all times I needed. I want to express my thanks to:

José Devezas, who always was and is not only my colleague but friend, and who helped me go through the difficult moments of my PhD journey, with good advice, new ideas and support.

João Castro, who was my supervisor at the beginning of my research and academic career, for patience, for good orientation in the scientific world and in Information Science.

Mafalda Lopes, who was my colleague in data management research interests and became a good friend, for her support, ideas, motivation, good food and always open doors.

Lígia Carvalho, who was and is my friend, given to me by my guardian angels, for her soul, inspiration, understanding, calmness, confidence, wonderful fests with great food and company.

Júlio Campos, who was and is my friend throughout my academic career, for his support and his ideas, for his family and his friendship.

Yulia Uskova, who was my English teacher and became my friend, for her patience, support, availability, time and dedication to improving my thesis and my English knowledge.

And other colleagues and researchers who turned into my friends during this wonderful time, who helped me with the development of my work, responded to surveys, and interviews, growing up together with me. Thank you very much!!

Also, thanks to several entities for providing essential conditions for the performance of my PhD thesis. As such, thank you:

* The Fundação para a Ciência e a Tecnologia, due to the grant that was provided to me, availability to clarify all issues related to the grant, and believing in me in general.

* To the Institute for Systems and Computer Engineering, Technology and Science for support on many issues, both technological and organizational, permission to consult the internal rules and processes to perform my work, many interesting case studies to collect the data, availability for interviews, meetings and for being the great host institute.

* To the Faculty of Engineering of the University of Porto, for total support and for good conditions to work.

Finally, I would like to express my thanks to all those who, in one way or another, made this PhD thesis possible.

My sincere and profound **Thank you all**!

SUPPORT FUNDING ACKNOWLEDGMENTS

The funding for this work was given by the PhD grant SFRH/BD/136332/2018, provided by the FCT - Fundação para a Ciência e a Tecnologia.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Nowadays, the value of research data is increasingly recognized. According to the Committee on Data for Science and Technology (CODATA)[1], the International Council for Science (ISCU)[2] and the Open Data Institute[3], the potential of open science is clear and contributes to social and economic growth, promoting auditability of results, reuse, and transparency [122]. Moreover, the 17 Sustainable Development Goals (SDGs)[4] adopted by the United Nations Member States in 2015 to end poverty, protect the planet, and enjoy peace and prosperity by 2030 require broad collaboration within the global data community, and the scientific community in general, as well as developing information infrastructures [214, 40, 282]. Achieving the SDGs is in alignment with best practices, recommendations, and Research Data Management (RDM) activities of the Research Data Alliance (RDA), one of the RDM initiatives for the Open Science paradigm [214].

At the same time, the research projects are characterized by a huge amount and a wide variety of data, which causes many issues related to data access, description, reuse, and RDM activities in general [266]. Good data management is critical to the quality of research data [89, 50] since data should be described with as much detail as possible to be Findable, Accessible, Interoperable and Reusable (FAIR principles) [89, 50, 162, 274].

In this context, the wide adoption of RDM best practices is an essential step toward data reuse. In Chapter 2, an overview of the FAIR principles and RDM best practices landscape are provided. Moreover, Section 2.4 of the same chapter presents the existing RDM initiatives with attention to the solutions proposed to support researchers in RDM activities that allow making the data more FAIR.

Despite the increasing interest in making research data available [258] and the existence of the institutional infrastructures and workflows designed to support researchers in RDM activities [89, 50], researchers still have to deal with several problems, such as the amount of time it takes and the inadequacy of the existing tools [236, 116, 228, 254]. Moreover, with the emergence of RDM requirements for these infrastructures and funders' requirements for grant applications, which are described in Chapter 2, the necessity of developing adequate and user-friendly tools or improving existing ones to meet all requirements is visible [48, 273].

Repository certification, described in Chapter 2, Section 2.6, also plays an important role nowadays because certified repositories are trustworthy, have useful RDM mechanisms, and are more recognizable in the research community. Thus, organizations are interested in the Core Trust Seal[5] certification of their repositories [67]. The selection of repositories almost always takes place at the project preparation stage, which is an important part of good planning. Moreover, researchers should specify the repository during the Data Management Plan (DMP) creation, which in many cases is also required when applying for a grant [45]. The Data Management Plan is a formal document that describes a strategy for research data management during the project, and even after its completion. It describes detailed information of the research data management activities and rules established in the project and is required by funders either along with the project proposal or within 6 months of its initiation. Moreover, it proves that the received funds "yield high-quality and reusable research data" [170].

1 http://www.codata.org
2 https://www.icsu.org/
3 https://theodi.org/
4 https://sustainabledevelopment.un.org/
5 https://www.coretrustseal.org/

A good DMP must include information about how data will be collected, processed, shared, and reused, how they will be described and where they will be preserved. Moreover, the context of the project, people in charge of RDM, and possible ethical and legal issues are also parts of a good plan [267, 241, 199]. A DMP can be regarded as a living document [6] that is useful for structuring the course of research activities, integrating with other systems and workflows [241], and encompassing the entire strategy of the project [55]. Chapter 3 of our work is devoted to issues related to the DMP, its goals, the importance of its creation and monitoring, its value and benefits.

To facilitate DMP creation, where possible, processes should be automated, making plans more useful and "live" [175]. To this end, in Chapter 3, Sections 3.2 and 3.3, we describe machine-actionable DMPs (maDMP)[7] and present initiatives, projects and tools developed to simplify and automate the DMP creation process and the exchange of information between stakeholders and RDM systems, which reduces the workload related to the DMP [241, 176].

In Chapter 3, Section 3.4, the Data Domain Protocol is described. It is proposed by Science Europe[8] and is supposed to promote standardized machine-actionable blocks for DMP [175]. Moreover, it can simplify DMP creation and its monitorization within specified disciplinary domains. In other words, the Domain Data Protocols can be seen as flexible metadata models with controlled vocabularies[9] for different scientific domains to simplify DMP-related processes.

Moreover, the Domain Data Protocols help to develop a standard for individual research domains, encouraging scientific communities to contribute to establishing protocols that meet their specific needs [73]. In Chapter 2, Section 2.5, we provide detailed information related to flexible metadata models and controlled vocabularies. Section 2.6 describes the Dendro platform on which controlled vocabularies have been implemented and the work done to improve metadata quality. The INESC TEC data repository development and methodology for data publication are described in Chapter 2, Section 2.6.

Some institutions already have well-established RDM workflows, while others are only beginning to develop RDM systems and tools. In general, in order to support their researchers, institutions require a great deal of interdepartmental communication and collaboration, as well as nationwide collaboration, to establish policies, practices, and regulations as required in the scientific community [8]. In this context, Chapter 6 describes the work and the collaboration with the different stakeholders required for the establishment of an institutional RDM workflow. This work includes an analysis of the state of the RDM workflow and its integration with different RDM tools described in Chapter 2, Section 2.7, analysis of the project management process at the institution (Chapter 6, Section 6.1), as well as the identification of the aspects that can be improved with the primary focus on the DMP support system (Chapter 6, Section 6.2, 6.3).

The creation of a DMP, the same way as most RDM activities, is a time-consuming task requiring some effort, specific knowledge, data publication experience and appropriate tools [228]. All RDM activities, and DMP creation in particular, are time consuming and therefore fail to motivate researchers. Therefore, many institutions are looking for solutions to help researchers perform RDM tasks by providing and developing tools, methods, support systems, and services, and by developing and implementing RDM workflows into the institutional infrastructure [39, 75, 279].

6 A living document is a document that, once created, needs to be up-to-date and reflect all changes related to research data management that have occurred during the project and even after its completion.

7 Machine actionable DMPs make text-based DMPs readable and usable by machines to connect all DMP-related resources, allowing research data and information to be communicated and shared across stakeholders, linking metadata, repositories, and institutions.

8 https://www.scienceeurope.org/

9 Controlled vocabularies are organized arrangement of concepts (words or phrases) used to index content and/or to retrieve content through browsing or searching. They typically include preferred and variant terms and have a defined scope or describe a specific domain [110].

Tools, support systems and services should assist researchers during the project lifecycle, clarify issues related to RDM issues and boost the visibility of research [238, 278, 153]. In this context, in Chapter 4, we present a systematic review that helped us to analyze the various existing ways to support DMP and their development, which organizations provide to help researchers. Another existing way to support researchers in creating a DMP is described in Chapter 5. This collaborative DMP-building method is the foundation of this work and contributes to the development of an institutional DMP support system, which is the main goal of this work.

The process of developing a DMP support system, its application, systematization, and testing at the institution with projects from different scientific domains, as well as its evaluation, results, discussion, conclusion, and future works, are described in Chapters 7 - 9.

1.1 GOALS

There is no established DMP support system which provides a solution to the full set of problems. The main objective of the work is to develop a DMP support system which can be implemented as part of the institutional RDM workflow, improving the effectiveness and efficiency of data management. It should assist researchers in DMP creation and monitoring, simplifying these processes, reducing the time spent and improving the quality and detail of the DMPs at the same time. Moreover, the DMP support system and data steward[10] support should help researchers in research data management in their projects, from the creation of the project proposal to its completion, bringing the data in line with RDM, funders' requirements and FAIR principles.

In this work we analyze existing RDM and funders' requirements, initiatives, infrastructure, tools, best practices and different institutional ways which aim to support researchers in RDM issues, including DMP creation. The DMP support system is based on the collaborative DMP-building method proposed for helping researchers in DMP creation and monitoring processes that are already used in the Institute for Systems and Computer Engineering, Technology and Science (INESC TEC) [11]. We also analyze RDM workflow and project management at INESC TEC and propose improvements according to the maDMP principles, automation mechanisms, flexible metadata models, controlled vocabularies, and Domain Data Protocols.

The DMP support system is developed, proposed, tested, and evaluated at INESC TEC, taking into account ongoing efforts in relation to DMP issues and existing RDM solutions in the research community proposed by significant RDM organizations. This work can be seen as a guide for institutions that are at the beginning of developing their RDM infrastructure and support services. Moreover, it can be used by any data steward to support researchers, using already proven methods and guidelines.

10 In our work, where I played the role of data steward at INESC TEC institution, the data steward is seen as responsible for implementing the RDM policies and recommendations defined by the Open Science paradigm, RDM and funder requirements, acting as a central element between researchers, institutional departments, and stakeholders involved in RDM. The data steward is involved in the organization of the RDM infrastructure and can support researchers in various RDM tasks. Moreover, the role requires specific knowledge, skills, and experience in data-related processes such as data deposit, description, organization, publishing or preservation.

11 https://www.inesctec.pt/en

1.2 RESEARCH QUESTIONS

In summary, this work is focused on the establishment of RDM workflow, the development of the DMP support system, and the proposal of the controlled vocabularies and Domain Data Protocols. The question of the importance and inclusion of certified repositories to be implemented in the RDM workflow is also included in this work. More specifically, the work is directed to the following research questions:

RQ1. How will a DMP support system be incorporated into institutional RDM workflow, becoming a required initial activity of each new research project?

RQ2. Can flexible metadata models be devised for each domain with a reasonable effort, taking advantage of the Domain Data Protocols defined by the Science Europe Working Group?

RQ3. Based on the preliminary examples of the DMP creation methodology, and the concept, and emerging standards for machine-actionable DMP, how can we devise DMP monitorization mechanisms that enforce compliance of the project results with the decisions taken during DMP creation?

RQ4. Will the proposed DMP support system satisfy the requirements of multiple scientific domains and the funders while keeping track of the data management effort in projects?

RQ5. Can the DMP support system include the selection of target repositories while offering an institutional solution certified with the Core Trust Seal?

1.3 METHODOLOGY

The progress of the work is based on contacts with INESC TEC researchers and researchers from other organizations (e.g., the University of Porto, Portuguese Oncology Institute of Porto (IPO-Porto), i3S - an institution devoted to research and innovation in the Health Sciences). We focus on the identification of their RDM needs with specific attention to DMP creation, supporting them during RDM tasks, and analyzing their internal project management procedures. Moreover, we study the regulations and requirements regarding RDM issues at the international and national levels, analyze existing RDM initiatives, tools, and infrastructures, and explore FAIR principles and funders' requirements. A systematic review of the different ways of institutional support for DMPs is conducted, and case studies of DMP creation and monitoring processes are followed and surveyed. In order to answer the research questions and achieve defined goals, a participatory-design research method [261] is used, involving researchers from different scientific domains.

This approach focused on collaboration with researchers to support them in creating, monitoring, or improving DMPs, beginning with exploring their scientific domains and the project context. To do this, we interview the researchers and analyze their needs, difficulties, and motivations for DMP creation. We also explore their publications and published data relevant to the project, if any. Information on selected data repositories, metadata standards used for description, and other important information related to RDM in the project is also analyzed. As a result of this preparatory phase, the necessary information is collected. It helps us not only to analyze the needs of researchers, but also to create the first version of the DMP for their project and, in some cases, to identify the terms used to develop domain-specific flexible metadata models. We motivate researchers to collaborate by explaining the value, benefits, and importance of creating a DMP, mentioning existing RDM commitments and funders' requirements, and encouraging them to publish their data.

Next, the internal INESC TEC procedures related to RDM and RDM workflow are analyzed. The role of the data steward at the institution is to interact with the project management teams and other institutional stakeholders, and organize inteviews and meetings. Their views on the importance and benefits of an improved

RDM workflow in the institution, as well as the need to develop a DMP support system, are also explored. This collaboration as a whole helps to define the lifecycle of research projects at INESC TEC with respect to data management.

We also review requirements and legislation related to research projects and data management at the international and national levels. This allows us to focus on the compliance of the projects with RDM and funders' requirements when creating plans.

Then, basing on the existing RDM workflow at INESC TEC, the processes that could be automated and simplified when creating the DMP are identified. Development of the DMP support system and its evaluation are proceeded.

The development of the DMP support system is based on the collaborative DMP-building method proposed in 2019[12]. In the course of the work, this collaborative method is analyzed, changed, and improved in order to be suitable for any type of project, in any scientific domain. To systematize and evaluate the collaborative DMP-building method, a sample of the projects is determined, and a series of collaborations with researchers from this sample is performed. The results of the collaborations are analyzed, and a list of recommendations for data stewards is created. Controlled vocabularies and examples of the Domain Data Protocol for DMP are also proposed, showing that by using them, the DMP support system can be more flexible, compatible, and machine-readable [73].

The collaborations are carried out in accordance with a defined sample of the different projects. Each of them is aimed at either creating a DMP for the respective project, or monitoring, or improving an existing one. As mentioned above, the DMP creation and monitoring processes are based on the collaborative DMP-building method currently used at INESC TEC and include tasks such as interviews, analysis of publications and data, and collection of other relevant information related to the domain and the research project.

The DMP creation also includes the analysis of issues related to the target data repository, namely its trustworthiness and certification, as well as support to researchers on data description and any issues related to RDM throughout the project's lifecycle and even after its completion. Considering the certification of repositories with the Core Trust Seal is a way to guarantee the quality, trust, recognition in the scientific community, and compliance with RDM requirements.

All work with the researchers is documented on the dataset "Assessment of metrics for the development of an institutional DMP support system" [131] to evaluate the collaborative DMP-building method, improve it, and propose a well-designed institutional DMP support system. We also analyze the effort and time required for DMP creation, the satisfaction, motivation, value, and importance of researchers for DMP support, the importance of having a data steward and DMP support system in the institution, and the difficulties the researchers face in DMP-related tasks. During the collaboration with researchers, we identify different support scenarios for creating and monitoring DMPs, which help to improve and expand the existing collaborative method. Institutional RDM policy, terms of use of the INESC TEC research data repository, and an institutional DMP support system are proposed too.

1.4 PRELIMINARY WORK AND CONTRIBUTIONS

One significant project aimed at developing RDM tools and analysing the needs of institutes and researchers is TAIL project - Research data management from creation to deposit and sharing (2016-2019) [13]. During this project, flexible metadata models and controlled vocabularies were developed for the Hydrogen Production scientific

12 https://www.rd-alliance.org/system/files/dmp_pia_method_yuliakarimova_41_.png
13 https://www.inesctec.pt/en/projects/tail

domain. Those controlled vocabularies allowed to make the Dendro platform [14] more flexible for use and suitable for multiple domains. Also, an evaluation of the use of the implemented controlled vocabularies showed that the process of data description required RDM skills, effort, time, and adequate tools, as the quality of the metadata could guarantee accuracy and access to digital resources, allowing others to find, understand and reuse them. As a result of this initiative, we observed that controlled vocabularies normalize the description, reduce syntactic and grammatical errors and obtain a more complete and correct description with similar effort. In other words, that work contributed to the development of the research data management tool from data production to deposit and share, namely the Dendro platform. Moreover, that work showed the difficulties that researchers face in the course of RDM activities and their need for support. The Dendro platform was one of the first tools added to the RDM workflow proposed for INESC TEC [133, 132].

The collaboration with the EUDAT[15], a European common data infrastructure with integrated services for data preservation and dissemination, allowed them to contribute to the development of their services and expand institutional RDM workflow, including B2NOTE for semantic annotation [151]. This collaboration was based on a Data Pilot established between the TAIL project and EUDAT to integrate services for the RDM workflow of the University of Porto [83] and showed that data reuse is strongly influenced by the quality of the metadata and detailed description of the datasets. Contribution from our side was focused on the test and evaluation of the EUDAT B2NOTE service[16], incorporated as part of the RDM workflow in INESC TEC. Its aim was to help researchers annotate datasets using a combination of semantic tags from controlled vocabularies, free-text keywords and free-text comments, making the data more understandable and reusable for others [137, 136].

The development and configuration of RDM tools allowing research data sharing and reusing, also had an important role in institutional RDM. In this context, the INESC TEC[17] research data repository was established. It was based on CKAN (Comprehensive Knowledge Archive Network)[18], an open-source data platform built as a data management system. The repository was incorporated into the RDM workflow of INESC TEC, and a methodology to support users in the data deposit and description was also created [135].

Ongoing collaboration with researchers from different domains and the emergence of the RDM requirements related to the submission of the Data Management Plan allowed us to develop the collaborative method for DMP creation between researchers and a data steward [146]. That collaborative method[19] was presented at the Research Data Alliance 14th Plenary in 2019 [20], a participation that benefited from an Early Career grant from the organisation. We showed the structure of the collaborative method and the beginning of its application at INESC TEC, the importance and value of creating DMPs and their monitoring. The collaborative method serves as a starting point for the development of the DMP support system described in this work, requiring appropriate improvements according to RDM and funders' requirements and its application in more projects for systematization.

The application of the collaborative DMP-building method and the creation of new tools and services in this area led us to the analysis of the maDMP standard proposed by the RDA DMP Common Standards Working Group[21], as well as the need to automate the DMP monitoring mechanism. While participating in the RDA Hackathon on maDMPs[22], our "InsTmaDMP" team[23] analyzed the existing RDM

14 https://github.com/feup-infolab/dendro
15 https://eudat.eu/
16 https://b2note.eudat.eu/
17 https://rdm.inesctec.pt/
18 https://ckan.org/
19 https://www.rd-alliance.org/system/files/dmp_pia_method_yuliakarimova_41_.png
20 https://www.rd-alliance.org/blogs/rda-fourteen-plenary-early-career-experience.html
21 https://tinyurl.com/mr2x6kcr
22 https://rda-dmp-common.github.io/hackathon-2020/
23 https://github.com/RDA-DMP-Common/hackathon-2020/blob/master/results.md#instmadmp

workflows at different institutions, their DMP creation process and compared them with the maDMP concept, identifying what should be changed to make the institutional DMP machine-actionable [141].

In addition, constant collaboration both nationally and internationally with different communities, initiatives, and RDA groups allowed us to contribute to various important RDM topics, such as metadata, controlled vocabularies, data citations, data repositories, and data management plans. A complete list of our contributions, which includes publications, posters, guidelines, workshops, training sessions, and published DMPs can be seen in the Appendix A. All the work done played an important role and influenced the development of the DMP support system.

1.5 SUMMARY

This introduction provides the context for the work, showing the importance of data sharing and reuse, policies and legislation established to promote the Open Science and Open Data movements in the scientific community. It also overviews the needs and challenges in developing RDM infrastructure, tools, services, and support that can help institutions and researchers meet all requirements related to data. In this chapter, we show that developing an RDM infrastructure involves many different aspects, such as compliance with FAIR principles, certification of repositories, data description, and metadata quality issues, DMP creation and monitoring processes, and other RDM activities that researchers should perform during their projects. Then, we point out that many institutions are trying to support researchers in different ways, developing support services, but researchers are still facing problems, more specifically with DMP creation. This introduction brings us to describe the thesis goals and research questions. We also describe the methodology used and give an overview of the previous work on which the developing DMP support system is based and the contribution that we are making to science.

2 RESEARCH DATA MANAGEMENT

Nowadays, we face times of adaptation, continuous learning, and innovation. With the development of new information technologies, new forms, and methods of scientific collaboration are emerging. In this context, research plays a fundamental role because without new projects, new discoveries, new data, knowledge, and improvements, we will stand still and stop evolving. The scientific community, government institutions and funders are interested in the advancement of scientific knowledge and promote it through initiatives related to Open Science and Data, data sharing, developing and improving services and tools to support RDM, and encouraging researchers to continue to contribute to society [163]. More and more scientific domains are becoming involved in data-driven science, where large amounts of data are created or reused to support research projects [105, 28, 27]. In this context, with the increasing amount of data being produced, RDM activities are getting justifiably recognition in order to stimulate the reuse of research data.

In this chapter we describe the emergence of the need for a good RDM, showing the most important aspects of developing this endeavor. Specifically, we consider what is meant by research data, what types of data exist, and what importance and value they hold for society in general. We also look at what policies, requirements, and overall work are being developed in the scientific community to create infrastructures, tools, and systems to help institutions, organizations, and researchers to bring their projects and outputs in line with scientific growth. Metadata, controlled vocabularies, and data repositories are also important for RDM because they help improve metadata quality, data access, and reuse. In this chapter, we explain how they can facilitate some of the RDM activities that researchers must perform in their projects to keep up with the demands of the scientific community and the evolution of science.

Finally, we describe different types of repositories and platforms for research data, the importance of their certification, and the example of the establishment of the institutional RDM workflow, integrating different RDM tools. Part of this work is based on published papers: Karimova, Y., et al. "Promoting semantic annotation of research data by their creators: a use case with B2NOTE at the end of the RDM workflow", 2017 [136]; Karimova, Y., et al. "Description + annotation: semantic data publication workflow with Dendro and B2NOTE", 2017 [137]; and Karimova Y., et al. "Data Deposit in a CKAN Repository: A Dublin Core-Based Simplified Workflow", 2019 [135].

2.1 RESEARCH DATA: VALUE AND IMPORTANCE

Due to the digital revolution, the quantity of data is growing every day, increasing its importance and value for society, for reuse, or for new research challenges [163, 105, 28, 27]. Data have a huge potential for improvements in the economy, politics, and citizen life [283]. Without data, it is not possible to validate project results, replicate findings, or use results as a basis for new developments [163]. Research data are valuable resources and usually require a lot of time, effort, and money to be produced [89]. Moreover, research data help to improve the reproducibility and transparency of scientific results and play a central role in the future of science [163].

The definition of data is complex because data can be categorized by forms, types, formats, purposes, or methods, among other characteristics [28, 55, 23]. However, there are more common definitions:

– *"Data are facts, numbers, letters, and symbols that describe an object, idea, condition, situation, or other factors"* [54];

– *"...recorded factual material commonly accepted in the scientific community as necessary to validate research findings..."* [1];

– *"Research data from public funding: factual records (such as numerical scores, textual records, images, and sounds) resulting from research that is partially or fully funded by public funds, used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings"* [2];

– *"..facts, observations, images, computer program results, recordings, measurements or experiences on which an argument, theory, test or hypothesis, or other research output is based"* [3].

In this context, we can state that all definitions of research data point to the outcome of research processes or activities which help to validate research findings and results.

Research data can be classified in various ways: by its nature (e.g., images, videos, etc.), by the discipline of origin, by the method of collection, according to the type of research (e.g., observational, computational or experimental) [23, 240, 28, 276, 288] and may be collected, generated, created, or reused during the research projects of all disciplines [163, 105, 109].

Observational data are data captured in real-time, usually unique and irreplaceable, cannot be collected again, and may involve specific places or times. So, they are usually preserved indefinitely (e.g., records related to the sea temperature on a specific date) [240, 50].

Computational data are the result of model execution or simulation on a computer and may include a description of the hardware, software, or other relevant specific information. Preservation of this type of data may not be necessary in some cases because they can be easily reproduced. In other cases, recreating or reusing the model requires a huge effort due to the description of specific information, documentation, hardware, etc. It is therefore recommended to preserve this type of data; sometimes models, sometimes outputs or documentation [240, 50].

Experimental data are mainly obtained in controlled environments, such as laboratories, and are usually reproducible. However, they can be very expensive to obtain, and the necessary experimental conditions are difficult to reproduce. For this reason, the long-term preservation of this type of data is also necessary [240, 50].

Along with observational, computational, and experimental data, the National Science Board (NSB) identifies one more type of data: records. Records are data obtained from social and humanistic research, such as public records of the government and public and private life [240]

Moreover, according to the report presented by the World Bank, data can be divided depending on two purposes: public (public intent data) or commercial (private intent data). Also, data can be divided by collection method: traditional or new, where the traditional method is related more to survey methods (e.g., census) and new - to digital tools (e.g., location data from satellite imaging). Both types of data, public and private, can be collected by traditional or new methods [283].

With establishment of the General Data Protection Regulation (GDPR) sensitive and personal data have been receiving more attention. Personal data are defined as "any information which are related to an identified or identifiable natural person", due to which directly or indirectly the person can be identified (e.g., telephone number, account data) [4]. Sensitive data are "data that can be used to identify an individual, species, object, process, or location that introduces a risk of discrimination, harm, or unwanted attention" (e.g., racial or ethnic origin) [103].

1 https://obamawhitehouse.archives.gov/omb/fedreg_a110-finalnotice
2 https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347
3 https://www.mopp.qut.edu.au/D/D_02_08.jsp
4 https://gdpr-info.eu/issues/personal-data/

Data can also be categorized according to their use (e.g., primary and secondary data use). Primary data are used for collection goals; secondary - for reuse or any use for goals different from the original [288]. Data can be quantitative and qualitative[5], and categorized by the data nature (e.g., images, videos, measurements, simulations, models, genomics sequencing, geospatial coordinates, drafts of manuscripts, notes, code, algorithms, among others) [288, 109, 55]. Some types of data have value immediately, others become more valuable over time [54].

In addition, research data can also be divided into different formats, such as **.pdf*, **.docs*, and others. However, the best practice accepted by the scientific community is to use recommended, open, standard, and common file formats for preservation. This is related to long-term preservation and data reuse, as some of the formats are more long-lived than others and do not require additional activities when reusing data [223].

Different types of projects with different scientific domains produce different types of data. Annotations on the lab notes, publications, protocols, and software, among others, are also considered as data. Collected and reusable data allow to produce new data upon previous work, continuing to develop initial ideas or create new ones. Sometimes, even the results of a project from one scientific domain could be reused as a beginning of a project from another scientific domain. This is why the data can be seen as "*multiple interdependencies*" of the research in general. In order not to break this chain, data [6] (i.e., dataset, data collection) should be opened, discovered, well described, easy to understand, interpret, and reuse [24].

In this context, the importance of the concept of Open Data, data management, data sharing and reuse has become a core focus of the scientific community, extending scientific knowledge, improving economic growth, and helping the public interest while reducing effort and costs for data reproduction or duplication [105, 163].

2.2 OPEN SCIENCE AND OPEN DATA

Currently, the Open Science and Open Data movements are promoted by government institutions, funders and journal publishers, increasing the value of the project results and maximizing investments [50, 283]. The importance of sharing research data becomes apparent during epidemics such as Ebola or COVID-19, where in a very short time period researchers need to act quickly to respond to public health emergencies and provide solutions to resolve the situation [50]. For example, during COVID-19 the unprecedented speed of vaccine development has been seen. Moreover, different apps have been developed to help monitor the spread of the virus or its effects. Thus, for this development, the data related to COVID-19 should have been shared and accessible [7].

There are different definitions of Open Science, for example:

– According to European Commission's Horizon 2020 programme, Open Science is *"the transformation, opening up and democratisation of science and research through ICT (Information and Communications Technology), with the objectives of making science more efficient, transparent and interdisciplinary, of changing the interaction between science and society, and of enabling broader societal impact and innovation"* [211];

– The United Nations Educational, Scientific and Cultural Organization (UNESCO) defined Open Science as *"an inclusive construct that combines various movements and practices aiming to make multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge*

5 https://eresearch.uwc.ac.za/research-data-management/research-data/
6 https://www.w3.org/TR/vocab-duv/
7 https://tinyurl.com/2ayyaywm

creation, evaluation and communication to societal actors beyond the traditional scientific community" [16].

Furthermore, according to the recommendation on Open Science created by UNESCO [8], the concept of openness makes science more transparent and accelerates the achievement of the Sustainable Development Goal [9] by connecting science, innovation and technology [16].

Overall, Open Science is a set of rules and practices that helps to share scientific knowledge, engage less-developed countries in knowledge sharing and improve access to technology development [16]. It aims to improve the quality of scientific knowledge, reproducibility, replicability, transparency, and credibility in science [157, 57, 25]. Moreover, it helps to increase the quality of the research [57] and it be seen as a return on investment [10].

The more data are open and available, the more research and development are possible, the more less-developed countries are able to participate in cutting-edge research, and the more opportunities for scientific progress emerge [28]. That is why, one of the main focuses is to develop policies which promote Open Science, invest in the infrastructures and services related to data openness, and promote international collaboration [16].

In this context, the Organisation for Economic Co-operation and Development (OECD) Principles and Guidelines for Access to Research Data from Public Funding [203, 190] and the Berlin Declaration on Open Access [17] can be seen as principal documents that have accelerated the Open Data movement in relation to research data, declaring data as a public good [50]. The European Commission also states that data are a valuable resource that could benefit society if the Open Science and Open Data movement goes alongside the development of the infrastructure [280, 42, 53]. The Recommendation on Access to Research Data from Public Funding [11] was revised and adopted by the OECD Council in 2021. The revision extends the existing version, highlights the importance of the openness of data during the COVID-19 pandemic, and reaffirms the importance of the openness and transparency of scientific results.

According to a report prepared by SPARC Europe [12] and Digital Curation Centre (DCC) [13], several EU Member States already have national policies related to Open Science [252]. Some of them are:

– Cyprus - National Policy of the Republic of Cyprus for Open Access to Scientific Information [14];

– Czech Republic - Czech Republic National Strategy of Open Access to Scientific Information [15];

– Spain - State Plan for Research, Development and Innovation 2017-2020 [16];

– France - National Plan for Open Science [17];

– Finland - Open Science and data-action Programme [18];

– Ireland - National Framework on the Transition to an Open Research Environment [19];

– Italy - National Plan for Open Science [20];

– Netherlands - National Programme Open Science [21];

8 https://www.unesco.org/en/natural-sciences/open-science
9 https://sdgs.un.org/goals
10 https://www.fosteropenscience.eu/content/winning-horizon-2020-open-science
11 https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347
12 https://sparceurope.org/
13 https://www.dcc.ac.uk/
14 https://tinyurl.com/zay8xryk
15 https://www.vyzkum.cz/Default.aspx?lang=en
16 https://www.ciencia.gob.es/Estrategias-y-Planes/Planes-y-programas/PEICTI.html
17 https://tinyurl.com/trmdnccx
18 https://edition.fi/tsv/catalog/book/64
19 https://tinyurl.com/n4yaky5z
20 https://tinyurl.com/dvw5pzde
21 https://www.dtls.nl/national-programme-open-science/

– Slovenia - National Strategy of Open Access to Scientific Publications and Research Data in Slovenia 2015-2020 [22];

– Slovak - Open Government Partnership National Action Plan of the Slovak Republic 2020-2021 [23].

In general, research projects should comply with the funder RDM requirements. For example, for projects funded by Horizon 2020, according to General Annex L of the Work Programme, it is recommended to become part of the Open Research Data Pilot and make data reusable and available to others [45]. In other words, researchers should deposit and share their data, making them reusable, with free access for any user. It is also emphasized that not all data can be open, so participation in the Pilot is not necessary but desirable and depends on the data collected/created during the project. The European Commission's approach, in this case, can be described as "*as open as possible, as closed as necessary*" [42]. The concept of openness can be applied to any scientific domain and includes such principal aspects as knowledge (e.g., publications, data, educational resources, code, etc.), infrastructure (e.g., tools, equipment, platforms, repositories, systems, etc.), communication, engagement, and connection with knowledge systems [16]. Moreover, data openness and data sharing can bring the following benefits:

– make new findings from old data;

– protect and preserve valuable resources;

– make results available to the public and to the scientific community;

– reproduce research and reduce the cost of reproduction or duplication;

– stimulate scientific research and development of science;

– promote innovation and potential of data;

– lead to new collaboration;

– maximize transparency in science and the responsibility of researchers;

– monitor and validate research results;

– increase the impact, value, and visibility of research;

– provide important and valuable resources for society, education, and training sessions;

– give researchers more visibility, credits, and citation in the scientific community;

– improve the economy, for example, by creating new posts of jobs or adding value to products and services [50, 89, 257].

In this context, the public and private sector, government institutions, funding agencies, national academies of science, research-performing institutions and scientific unions should be responsible for implementing Open Science policies and related requirements. They should encourage, promote, and invest in the development of policies, infrastructures, tools, and services that will help researchers do their projects and results in accordance with the requirements and principle of "*open by default*" [16, 252]. Moreover, research data should be placed in trusted data repositories. Publications should include citations to datasets with the results created during the project and other information that can help validate or reuse the research findings. If the data cannot be opened, justification is required; otherwise, the publication may be rejected [190].

All these aspects related to research data and their openness led the community to highlight the importance of the RDM activities, establish RDM requirements and define FAIR principles, which were presented in 2016 by a consortium of scientists and organizations [274].

2.2.1 Portugal Open Science policies

Regarding Portugal, the National Policy for Open Science is under development by the Government and the Ministry for Science, Technology and Higher Education.

22 https://tinyurl.com/2p88h24h
23 https://tinyurl.com/2p8v43ah

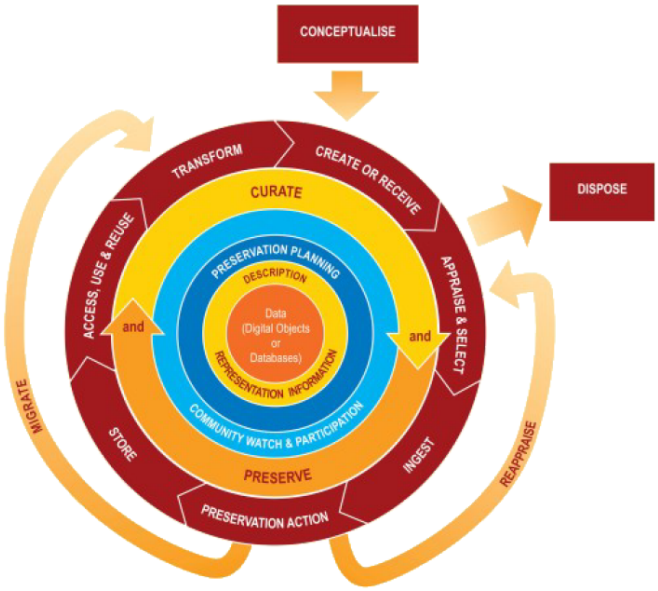

Figure 1: The DCC Curation Lifecycle Model [118]

The Portuguese National Funding Agency for Science, Research and Technology (FCT) already has a policy on data management and sharing [24] and a website dedicated to Open Science in the country [25], which indicates the main areas of work: transparency in practices, methodology, observation, and data collection; public availability and reuse of scientific data, public access and transparency of scientific communication; and use of web tools to facilitate scientific collaboration. This is not yet a mandatory policy, it is a recommendation document for FCT-funded research projects, encouraging researchers to open and share data whenever possible and to indicate the FCT as a funder of any dataset [252].

2.3 RESEARCH DATA LIFECYCLE AND RDM PRINCIPLES

2.3.1 Research Data Lifecycle

Data sharing only helps to increase scientific knowledge and benefits society if the data are available, preserved, understandable and reusable [148]. This is why data curation and management during their lifecycle are playing an important role in science now.

According to the DCC, which has played a leading role in the scientific community since 2004, digital curation requires effort, investment of time and appropriate resources. The DCC is one of the organizations that develop best practices, standards, and tools, collaborate with other RDM initiatives to support data creators and promote data sharing policies. The DCC presents a digital curation lifecycle mode (Figure 1) that defines a sequence and the correct order of components and activities, which together create a range of data-related processes. In other words, it represents data curation and the processes that go into RDM [163].

This model can be used in any scientific domain for any research project and includes activities such as data creation, processing, description, preservation, and data reusing, among others. In other words, this model describes full, sequential, and occasional data-related activities which can be seen as a basis for RDM [105,

24 https://www.fct.pt/documentos/PoliticaAcessoAberto_Dados.pdf
25 https://www.ciencia-aberta.pt/

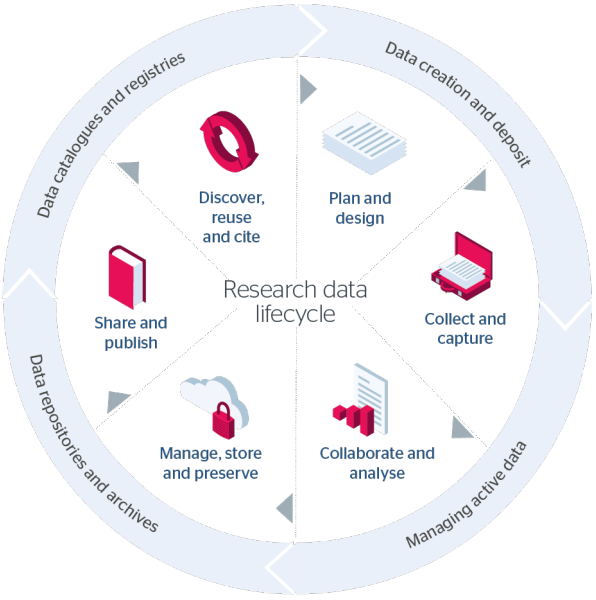

Figure 2: Research data lifecycle presented by JISC [124]

118]. The full activities refer to the description of data, representation of information, preservation planning, community participation and awareness. The sequential - cover data creation, conceptualization, evaluation, and selection for long-term preservation, access, reuse and transformation. The occasional - include elimination, re-evaluation and migration [118].

The lifecycle model can help to analyze the relationship between each stage and a data-related process, to understand the needs and lacks which exist in these stages, and to provide the information to identify data services or tools which should be developed [163]. The lifecycle can also help to understand what role data plays in each stage of the project [28] and realize how to manage, collect, process, preserve, and maintain the data [105]. Moreover, it will help to identify the key stakeholders in the scientific community who should be involved in all data-related activities, developing policies, tools, and infrastructure that will help at all stages of the project [61]. This model is only a suggestion and does not oblige organizations to fulfill the cycle from this first step, it can be changed depending on the needs of each organization [163].

Another example of the research data lifecycle has been proposed by the Joint Information Systems Committee (JISC) (Figure 2) in the JISC Research Management Toolkit [124], which was developed to support researchers in performing the data-related activities. This lifecycle includes different stages: the planning and designing, the data collection and description, data analysis, management, storage, and preservation, the sharing and publication stages, as well as discovery, reuse and citation [277].

There are other lifecycle models, but in general, all lifecycle models and their processes are related to data creation, processing, analysis, storage, sharing, use and preservation, where data reuse occurs until they are destroyed [283, 206].

2.3.2 RDM principles

Good data management is fundamental to the quality of research data. If data are well organized, documented, stored, and accessible, the result is better data management, which contributes to improved data quality [89, 27]. RDM includes activities and processes associated with the data lifecycle. In other words, RDM refers to the actions required to maintain data throughout their lifecycle and overtime for current generations and future uses [154].

There are different definitions of the RDM, for example:

According to European Commission, RDM is *"the process within the research life-cycle that includes the data collection or acquisition, organization, curation, storage, (long-term) preservation, security, quality assurance, allocation of persistent identifiers (PIDs), provision of metadata in line with disciplinary requirements, licencing, and rules and procedures for sharing of data"* [84];

The National Institutes of Health defines data management as *"the process of validating, organizing, protecting, maintaining, and processing scientific data to ensure the accessibility, reliability, and quality of the scientific data for its users"* [26].

RDM refers to *"the processes involved in organizing, storing, and sharing the data that is generated through the course of a research project"* [109] and is seen as an essential element in any project that helps to generate, collect or reuse data [84].

Overall, RDM is a set of activities that aims to ensure data reproducibility, prevent data loss, help to make data in compliance with requirements and includes the following stages:

– data planning;
– data collection;
– data organization and documenting;
– data analysis;
– data processing;
– data storage and preservation;
– data publication and sharing [288, 28, 55, 50].

RDM should be started at the planning stage of the project, carried out during the project and continued after the completion of the project [28].

Good data management should be organized both at the national and international levels to guarantee effective data reuse. Moreover, RDM policies should cover infrastructure policies (e.g., development of technical standards), laws and regulations (e.g., interoperability standards), economic policies (e.g., data-enabled services) and institutions (e.g., support for data management) [283]. All these actions should aim to provide researchers with the necessary RDM infrastructure and support services [190].

Along with the growing importance of research data and RDM, the scientific community is developing various requirements, policies, and principles. The implementation of RDM requirements is related to the funders' investments because they want to know how resources are spent by researchers and want to prevent data loss after the project is completed [266]. Funders encourage researchers to share their data and perform RDM tasks. Data reuse also helps to save funders' money because, more often than not, data collection is more expensive than data reuse [28].

2.4 RDM REQUIREMENTS, INITIATIVES AND INFRASTRUCTURE

Data openness helps the EU to be more competitive, improve products and services, as well as make data more valuable for both the economy and society. In this context, various policies, requirements, and strategies have been developed, adopted, and implemented in the scientific community to achieve all the goals related to sharing and reusing data [46]. All of these policies cover data-related issues ranging from researcher training to infrastructure development and emphasize the importance of RDM. Moreover, these policies have influenced the requirements for RDM [105, 50].

Over the past 15 years, the United Kingdom, United States of America, Australia, and EU have adopted Open Data and Open Science policies, focusing on RDM principles and promoting data sharing and reusing [105, 50, 190].

26 https://sharing.nih.gov/data-management-and-sharing-policy/data-management

A pivotal moment in the UK was the recognition and adoption of the Common Data Policy by the Research Councils in 2011. Various policies and requirements, such as the Concordat on Open Research Data, were then proposed, engaging UK's research funders (e.g., Biotechnology and Biological Sciences Research Council, Science and Technology Facilities Council) [105, 50, 190, 263].

In the USA, National Science Foundation (NSF) and National Institutes of Health (NIH) policies have played a fundamental role in improving data access, sharing, and reuse. Besides the common data-related aspects, they have also focused on requiring Data Management Plans which should include important information about data management for all grant proposals and on developing data infrastructure. The adoption of these policies has led to the involvement of research funding in the US, such as the US Department of Education, the US Department of Transportation, among others [50].

One of the main documents describing data openness and their management in Australia is the Australian Code for the Responsible Conduct of Research. It focuses on developing a framework for the responsible conduct of research, emphasizing the need for reliability and validity of research. It also highlights the importance of good data management practices, the need to comply with legislation and regulations, and the development of an appropriate infrastructure to help researchers meet all requirements [113].

In Europe, in turn, apart from the Data Open policies already mentioned, the main documents describing RDM requirements have been developed by the European Committee within the Horizon Europe work Programme. This document mentions RDM as mandatory. Researchers must focus on openness and reuse of data, provide a DMP within the first six months, make the data FAIR and ensure open access to their results. Regarding repositories, researchers could request the use of the European Open Science Cloud federated repositories [27], developed in Europe to support researchers in RDM activities [84].

Along with funders, journals, and publishers, such as Elsevier [28], Pangae [29], Nature [30] or PLoS [31], have also begun to adopt data sharing policies and require data submission along with publications [23, 50].

RDM requirements are varied, but in general they are all directed to data reusing and obtaining more benefits for society. In general, they include the following aspects:

– application of best practices and procedures related to RDM, with a focus on data sharing, reuse, and long-term preservation;

– handling data in accordance with the Open Data principle "*as open as possible, as closed as necessary*", relevant legislation and regulations regarding data use;

– justification with weightily reasons if data cannot be opened and shared;

– compliance with legal, ethical, intellectual property rights and confidentiality obligations;

– handling data in accordance with FAIR principles;

– provision of rich metadata;

– depositing data in a trusted repository whenever possible, with persistent identifiers and data citations;

– managing data during the project from the planning stage to completion;

– DMP creation and its monitoring [249, 52, 126, 8].

As requirements change rapidly, researchers are advised to check the RDM requirements for each funder before submitting an application. The new project applications for funding require well-planned RDM tasks described in the DMP. A good RDM can guarantee better data quality and easier reuse. So, researchers should make an effort to understand all RDM requirements in order to perform

27 https://eosc-portal.eu/
28 https://www.elsevier.com/
29 https://www.pangaea.de/
30 https://www.nature.com/
31 https://plos.org/

data-related tasks and make their project results more valuable to the community. Institutions, in this case, should provide support to researchers in all RDM activities [50].

Thus, in response to Open Science and Open Data policies as well as RDM requirements, educational and research institutions have also started to implement their own data management policies and develop services to support researchers/students to perform RDM activities. These policies vary from institution to institution. Some institutions are at an early stage of their development; others already have a well-established RDM infrastructure. Some institutions develop RDM services, tools, and infrastructure according to their own needs; others adopt the best practices existing in the scientific community. However, all institutional policies are based on legislation, policies, requirements, and regulations established by government and funding agencies and aim to share and reuse data [288].

2.4.1 FAIR principles

One of the RDM requirements is that the data should comply with FAIR principles. This is a requirement that should be fulfilled for any data in the project (totally open or with some restrictions). In other words, data should be FAIR, even if only the metadata describing the data are accessible. FAIR is not a synonym of Open Data, but it promotes data discovery, dissemination, and reuse, thereby accelerating scientific advances.

The FAIR principles were presented in 2016 by the Future of Research Communications and e-Scholarship Group (FORCE11) [32], which aims to join librarians, archivists, and funders together to improve and share knowledge [274].

The FAIR principles coincide with the research data lifecycle and also promote long-term validity and usability [50].

So, FAIR data mean:

– *Findable*. It is about making it easier for others to find the data by providing a description of the data, using rich metadata with a unique and persistent identifier (e.g., DOI), and registering or indexing data in a searchable resource (e.g., trusted repository).

– *Accessible*. This means that anyone can access data or metadata, using an identifier and an open, free and standardized communication protocol that allows appropriate authentication and authorization.

– *Interoperable*. It is about using formal, accessible, shared and broadly applicable language or vocabularies that conform to FAIR principles for data and metadata.

– *Reusable*. This means that the data should be comprehensible for reuse for anyone, have rich metadata, detailed provenance and a clear and accessible license (e.g. Creative Commons Attribution International Public License (CC BY)) [274, 16].

These principles are interlinked and aim to benefit data openness, making data easily to find, accessible, interoperable, and reusable, with a focus on reusing data by third parties [274].

The FAIR principles can be seen as guidelines for researchers, publishers, and organizations and help increase the reusability of scientific data. Data should be FAIR not only for humans but also for machines, allowing automated data retrieval and access [50].

Over time, the entire scientific community has followed these principles, turning them into a standard [50]. Thus, the Science Ministers of the Group of Eight (G8) [33] in 2013 [34] and later, in 2016, of the Group of Twenty (G20) [35] stated the importance of research data openness. Moreover, they affirmed the FAIR principles as appro-

32 https://force11.org/
33 https://en.wikipedia.org/wiki/G8
34 https://tinyurl.com/5n7t6554
35 https://g20.org/

priate, which should have helped make the data more available, accessible, and reusable by others. All of this has resulted in government institutions, international organizations, and funding agencies have also started to pay more attention to data openness and FAIR principles in the research community [50].

According to the report "Turning FAIR into reality" presented by the European Commission Expert Group on FAIR Data, FAIR principles have become very important in the scientific community [41]. The European Commission recommends implementing and supporting FAIR in research, whether it is infrastructure, technology, and tools, or legislation level. Moreover, they encourage the research community to develop the components of a FAIR ecosystem, promote, and reward FAIR sharing within disciplines and for interdisciplinary research, incentive training sessions, develop metrics to certify FAIR services, and provide strategic plans to achieve all defined goals related to FAIRness [41]. The European Commission adopted these principles along with an Open Data policy, making them guiding principles for data management planning, proposing them in the guidelines for Horizon2020 to promote the wide reuse of research data [84, 50].

This definition of FAIR principles has been placed in all Open Science and Open Data requirements and recommendations. One such example is the revised Recommendation on Access to Research Data from Public Funding of the OECD Council that in 2021 recognized the importance of FAIR data, where FAIR principles have been adopted and recognized as a necessity for *"enhanced access to research data"* [190].

Data, software, or other research resources and outputs are digital objects to which FAIR principles also should be applied. As well as data, software needs to be open to ensure transparency, reproducibility, and reusability, benefiting the economy, science, and society as a whole [150]. In this context, the RDA FAIR Principles for Research Software Working Group [36] has been organized to analyze and understand specific characteristics of software for proper application of the principles [150].

Various policies related to Open Science, Open Data, FAIR principles, and RDM can be found in a catalog that has been developed by the FAIRsharing community, which is based at the University of Oxford. This initiative aims to provide and promote information that may be useful to researchers in data-related processes and RDM issues, such as standards, policies, repositories, etc [227].

There are other FAIR-related initiatives and groups which attend to the growing importance of adopting FAIR principles. For example, the RDA FAIR Data Maturity Model WG [37] aims to develop recommendations on core assessment criteria for FAIRness; the RDA FAIRsharing Registry: Connecting data policies, standards and databases. RDA WG [38] aims to promote the adoption and improvement of FAIRsharing policies; the FAIRsFAIR (Fostering Fair Data Practices in Europe) initiative [39] aims to develop best practices in applying the principles throughout the research data lifecycle, emphasizing the importance of FAIR culture. Recognizing that RDM requirements and FAIR principles have become a key factor in the research community, the FAIRsFAIR initiative has produced training sessions and education guidelines for higher education institutions. The implementation of these practices is strategic in the institutional plans of universities and aims to stimulate Open Science in the community [80].

2.4.2 RDA - Research Data Alliance

Since 2006, many different initiatives related to Open Data, Open Science, and RDM have been launched [50].

36 https://www.rd-alliance.org/groups/fair-research-software-fair4rs-wg
37 https://www.rd-alliance.org/groups/fair-data-maturity-model-wg
38 https://tinyurl.com/bd56rxnw
39 https://www.fairsfair.eu/

One of the principal initiatives is the Research Data Alliance [40]. This initiative was organized by the European Commission together with the United States Government's National Science Foundation and National Institute of Standards and Technology, and the Australian Government's Department of Innovation and launched in 2013. The main aim of the initiative is to bring together professionals, researchers, and stakeholders, among others, to build social and technical bridges for data exchange and reuse. Professionals, researchers and innovators from all over the world with different backgrounds and scientific domains join the RDA and its Working Groups, and Interest Groups or even create national nodes to share knowledge, develop tools to solve RDM challenges, and influence local and regional governance [19, 18].

In Portugal, for example, an RDA-Portugal node [41] was created in 2018, led by a group of 8 institutions, such as the University of Porto, University of Évora, University of Coimbra, Instituto Superior Técnico of the University of Lisbon, Polytechnic Institute of Portalegre, Institute of Social Sciences of the University of Lisbon, University of Minho and Portuguese National Archives. As well as other nodes, this group is working on solving data sharing problems in their country, implementing and adopting existing RDM practices, developing tools and infrastructure for RDM in institutions and universities, analyzing the difficulties existing in the scientific community in Portugal, and promoting awareness about data sharing and the concept of Open Data [42].

In general, RDA hosts various types of events such as meetings, conferences, workshops, and training sessions where all interested parties can gain, or exchange their knowledge, or clarify questions related to data sharing issues and RDM requirements.

RDA currently has more than 12,600 members from 145 countries, 45 Working Groups and 60 Interest Groups whose topics include DMPs, data repositories certification, interoperability, metadata schemas and standards, and other educational, training sessions, and technology aspects related to data, FAIR principles, RDM infrastructure development, and implementation.

2.4.3 EOSC - European Open Science Cloud

In response to the requirements of Open Data, Open Science, and RDM, the European Commission has started to develop tools and infrastructure for research data [85].

In this context, the European Open Science Cloud (EOSC) [43], an environment to host and process research data to support EU science in line with RDM and FAIR requirements, was introduced in 2016 [44].

The main goal of EOSC is to provide an interdisciplinary environment and tools where researchers, innovators, data creators, and other interested parties can publish and find data, tools, guidelines, and services for data exchange and reuse [45].

There are different EOSC projects aimed at developing, coordinating, and establishing data infrastructures and services. Some of them are: the EOSC-Nordic project [46] which involves Finland, Sweden, Norway, Denmark, Iceland, Estonia, Latvia, and Lithuania; the EOSC-Pillar project [47] which includes Austria, Belgium, France, Germany, and Italy; the EOSCT-Synergy project [48] brings together Spain, Portugal, the UK, Czech Republic, Slovakia, Poland, the Netherlands, and Germany;

40 https://www.rd-alliance.org/
41 https://www.rd-alliance.org/groups/rda-portugal
42 https://tinyurl.com/aw7fx83k
43 https://eosc-portal.eu/about/eosc
44 https://tinyurl.com/4kds26mr
45 https://tinyurl.com/4kds26mr
46 https://www.eosc-nordic.eu/
47 https://www.eosc-pillar.eu/
48 https://www.eosc-synergy.eu/

and the NI4OS30 [49], which includes Greece, Cyprus, Bulgaria, Croatia, Hungary, Romania, Slovenia, Serbia, Albania, Bosnia-Herzegovina, North Macedonia, Montenegro, Armenia, Georgia, and Republic of Moldova [84]. The EOSC also brings together institutional, national, and European initiatives [50], creates a trusted virtual space, and engages all relevant stakeholders in data sharing issues [84].

This environment is developing in line with established policies and promoting FAIR principles in Europe. They already have an EOSC portal [51] where it is possible to find the latest updates regarding legislation, policies, project information, funding opportunities, and other relevant documents related to data sharing and reuse. In addition, there is an EOSC Marketplace, which provides various services and resources for researchers, and is classified by different scientific domains and categories [52].

Eventually, when fully developed, the EOSC will provide a Web of FAIR data environment with a wide range of services related to data sharing and reuse.

2.4.4 OpenAIRE - Open Access Infrastructure for Research in Europe

The Open Access Infrastructure for Research in Europe (OpenAIRE) initiative [53] was established in 2009. It is a network of experts and specialists related to Open Science and Open Data issues. Moreover, OpenAIRE is a large technical infrastructure that includes various services, documentation, training sessions, support, and knowledge pooling aimed at providing RDM services and promoting FAIR principles.

The same way as other initiatives, OpenAIRE cooperates with RDM stakeholders, institutions, and other initiatives, contributing to promoting the establishment of a culture of data sharing and reuse in Europe. Currently, there are 36 OpenAIRE's National Open Access Desks from the different EU countries, including Portugal and the Member States, with over 50 institutions working together to improve and implement effective Open Science and Data policies to raise awareness about data sharing and reuse at the national and international level, and to provide adequate tools to support researchers in all RDM activities.

Their current catalog of services [54] includes services for all stages of the research data lifecycle. For example, for the planning stage and creation of the DMP, they provide the Argos tool [55] which helps researchers to create plans; for data anonymization, researchers can use the Amnesia tool [56]; the Metadata Validator tool [57] can be used for verification of repositories' compliance and may be useful for repository managers. Together with the European Organization for Nuclear Research (CERN) [58], they have also developed the Zenodo repository [59], which is the main generic repository where researchers from different scientific domains can deposit and share their projects' outputs.

"We make Open Science happen" - is one of OpenAIRE's mottos.

2.4.5 DANS - Data Archiving and Networked Services

In 2005, Data Archiving and Networked Services (DANS) [60] started to provide a range of services related to data sharing and reuse, and help researchers make their

49 https://ni4os.eu
50 https://eosc-portal.eu/about/eosc-projects
51 https://eosc-portal.eu/
52 https://marketplace.eosc-portal.eu/
53 https://www.openaire.eu/
54 http://catalogue.openaire.eu/search
55 https://argos.openaire.eu/splash/
56 https://amnesia.openaire.eu/
57 https://provide.openaire.eu/home
58 https://home.cern/
59 https://zenodo.org/
60 https://dans.knaw.nl/en/about/

data FAIR, valuable, and reproducible. DANS was organized with the support of the Royal Netherlands Academy of Arts [61] and Sciences and the Dutch Research Council [62]. It is the Dutch national research data center, which provides various types of RDM support and certified services for short and long-term data preservation [63].

Furthermore, DANS provides a CoreTrustSeal [64] trustworthy data repository for research data from scientific domains such as social sciences and humanities, archaeology, life, health and medical sciences, and physical and technical sciences. This repository is one of the leading repositories in Europe.

DANS collaborates with other initiatives related to RDM issues. For example, DANS joined the EOSC association in 2020 to contribute to infrastructure development together with the EOSC Task Forces.

Regarding FAIR, DANS has developed two tools aimed at the assessment of data FAIRness [65]. One is the SATIFYD tool [65] that helps researchers or other interested parties to improve FAIR aspects of their data by providing FAIR analysis and advice on how to improve FAIR. Another tool, FAIR-Aware [66], helps researchers evaluate their knowledge related to the FAIR principles and understand the benefits of making their research data FAIR.

Their current and future directions are based on three pillars: expanding the services of the center of expertise for FAIR research data; continuing to improve the data repository in line with ongoing changes in data and their dimensions; increasing investment in active collaboration and coordination of both researchers and institutions as well as service providers and stakeholders [67].

2.4.6 ANDS - Australian National Data Service

Since 2008, the Australian National Data Service (ANDS), now renamed the Australian Research Data Commons (ARDC) [68], has begun supporting researchers and institutions in all RDM activities such as data organization, deposit, DMP creation, backups, long-term preservation, among others. This digital research infrastructure is a result of the collaboration between Monash University [69], the Australian National University [70] and the Commonwealth Scientific and Industrial Research Organization [71]. It is currently the main body concerned with the sharing and reuse of research data.

The ARDC focuses on data-driven science, improving the quality of research, data preserving, and sharing of high-quality information resources. The ARDC also develops various types of guidelines, training sessions, workshops, and events to help researchers ensure that their projects and outputs meet all RDM and FAIR requirements [288, 6].

The same way as DANS, ARDC also develops various types of services and tools. For example, they have developed the FAIR self-assessment tool [72] which allows researchers to analyze their data according to FAIR principles.

Another tool that distinguishes the ADRC from other initiatives has to deal with controlled vocabularies. The Research Vocabularies Australia [73] is a tool for researchers to create, manage, find, and access different vocabularies used in differ-

61 https://www.knaw.nl/en
62 https://www.nwo.nl/en
63 https://dlmforum.typepad.com/DANS_a_new_data_initiative_in_the_Netherlands.pdf
64 https://www.coretrustseal.org/
65 https://satifyd.dans.knaw.nl/
66 https://fairaware.dans.knaw.nl/
67 https://dans.knaw.nl/en/about/strategy/
68 https://ardc.edu.au/
69 https://www.monash.edu/
70 https://www.anu.edu.au/
71 https://www.csiro.au/
72 https://ardc.edu.au/resource/fair-data-self-assessment-tool/
73 https://ardc.edu.au/services/research-vocabularies-australia/

ent scientific domains. Furthermore, researchers can offer their own vocabularies and integrate controlled vocabularies into the developed system using the widget service [74].

2.5 METADATA, FLEXIBLE METADATA MODELS AND CONTROLLED VOCABULARIES

Since data frequently changes during the lifecycle of a project, a dataset for publication may contain different types of data collected or processed, as chosen by the researchers [163]. Therefore, a description of the datasets and their metadata can help in understanding the whole picture of the study, its specificity, methodology, and the tools used. Therefore, metadata nowadays plays an important role in data-driven science [274, 16, 41], and are not only important for humans but also for machines and information technology, such as search engines, tools, systems, digital libraries, and museums - anything that involves data [106, 281].

Good and quality metadata depend on researchers. They must provide accurate and complete descriptions and rich metadata to facilitate reuse of the data and to communicate their project ideas and results to others [163]. In other words, metadata help to make data more understandable and correctly interpreted, both by humans (e.g., specific weather conditions at the time of data collection) and machines (e.g., description of users used to authenticate to the system, file creator, access time, usage statistics) [276, 284]. Without metadata, it is difficult to understand what data are collected, what they represent, and whether there are any specifics for reuse. Also, in some cases, data without metadata become an ordinary set of digits and symbols, unusable and understandable by others [249, 79].

In general, metadata are associated with transparency in science, making data easier to find, understand, and reuse. Transparency increases when the quality and intensity of metadata increases too [79].

The first mention of metadata as "meta language" comes from MIT's Stuart McIntosh and David Griffel, in 1967. In 1968 Philip Bagley used the term "metadata" in the ISO 11179 as "data about the containers of data" and since then the concept "metadata - data about data" has been adopted by scientific communities, including computer science, library and information science [75]. Metadata have been classified as "*the backbone of digital curation*" [107], because they allow the identification, discovery, understanding, preservation and use of digital objects [76]. In general, metadata are defined as "*data about data*", structured data about data [77], or information about information [222, 107].

There are other definitions of metadata. However, they all refer to the ability to make data and digital resources more accessible, manageable, understandable, and reusable [272].

From a global perspective, metadata are descriptive or contextual information associated with a digital object or resource. Metadata can be categorized according to the purpose of the description, whether it is a more technical description including, for example, file size and format or an administrative description describing processes applied over time [222]. In addition, metadata can be descriptive, structural, and for preservation, which allow location, identification, and connection to other digital objects [210, 281]. Metadata can also be considered operational, providing information automatically generated by the system, such as date or timestamp [288].They can also provide information about property rights and other important specific information that the data creator wants to add to simplify

74 https://vocabs.ardc.edu.au/vocabs/page/widget_explorer
75 https://en.wikipedia.org/wiki/Metadata
76 https://tinyurl.com/ycy68fz8
77 https://tinyurl.com/2t7u97xd

data reuse. In this context, metadata can be seen as the basis of the FAIR principles, because each FAIR principle describes the rules for dealing with data and metadata. In other words, metadata refers to the descriptive process in RDM and helps to make the data more FAIR [281, 274, 16, 41].

Metadata are as important as data, and their creation takes time and effort. For this reason, various directions have been created in recent decades to facilitate metadata creation and improve their quality. One of these directions is related to the development of metadata schemas that can standardize and facilitate data description [276].

2.5.1 Metadata schemas, generic standards and flexible metadata models

Research data vary greatly in type. So, these differences are reflected in metadata schemas and standards, which support data reuse and managing. [276].

In the context of metadata, the term schema defines admissible elements and values (name-value principle), the format they should take, and in some cases the maximum and minimum lengths. The main purpose of schemas is to facilitate the automatic exchange, interoperability, and processing of information from different sources (e.g., in repositories or other information systems). Semantics, content rules, and syntax are three aspects of metadata that can be specified in metadata schemas [32, 180, 200].

In recent years, many different metadata schemas and standards have been developed. The Dublin Core standard [78] has been one of the principal and more widely established in the computer science and library, and information science community. It focuses on describing information resources and is easy to use for anyone. This standard is developed by the Dublin Core Metadata Initiative, documented by ISO 15836:2009 [79] and contains fifteen basic schema elements. They are: Contributor, Coverage, Creator, Date, Description, Format, Identifier, Language, Publisher, Relation, Rights, Source, Subject, Title and Type [200].

There are more common standards such as the MARC (Machine Readable Cataloging) for bibliographic items [80], theEAD (Encoding for Archival Data) for a description of archival collections [81], the CDWA (Categories for the Description of Works of Art) for art objects [82], the VRA Core (Visual Resources Association Core) for visual works of art and architecture [83], among others [11]. Some of them are for cataloguing library documents, and others have been created to find electronic resources in the web environment.

A large list of standards can be found in the Metadata Standard Catalog [84] created by RDA Metadata Standard Directory Working Group [85] which in turn is based on the list created by the DCC [86]. This catalog includes different standards from different scientific domains, such as Social Sciences, Humanities, and Atmospheric sciences, among others.

In addition to generic metadata standards, there are also domain-specific ones. Although such metadata are more time-consuming to describe than general metadata, they do contribute to deeper and more precise descriptions and enhance the quality of domain-specific metadata [38].

In this regard, many disciplines have developed their own specific standards [38]. For example, the Ecological Metadata Language [87] for describing biological and

78 https://www.dublincore.org/specifications/dublin-core/dces/
79 https://www.iso.org/standard/52142.html
80 https://www.loc.gov/marc/umb/
81 https://www.loc.gov/ead/
82 https://www.getty.edu/research/publications/electronic_publications/cdwa/
83 https://www.loc.gov/standards/vracore/
84 https://rdamsc.bath.ac.uk/
85 https://www.rd-alliance.org/groups/metadata-standards-directory-working-group.html
86 www.dcc.ac.uk/resources/metadata-standards
87 https://eml.ecoinformatics.org

ecological collections, the Darwin Core standard used by the biological community, and the Data Documentation Initiative (DDI) [88] for describing data related to social, behavioral, and economic research projects [276].

Although there are many different schemes and standards, all of them have common requirements (e.g., the schema must be extensible, flexible, understandable, and simple) and share the same goal, namely to support data use, reuse, and preservation [276].

No single metadata standard, neither generic nor domain-specific, can fully satisfy all researchers' needs. Therefore, to further improve metadata quality, application profiles and flexible metadata models based on the "*mixing and matching*" approach are proposed. These are combinations of elements (i.e., descriptors) from different domains and schemas, creating the best option to meet the descriptive requirements of researchers [115, 77]. Furthermore, the flexibility in the use of such models allows a better reflection of the context of the resource and its specificity, making it more interoperable [276]. In some cases where elements for a specific description are not available, a specific descriptor can be created together with vocabularies [38, 132].

2.5.2 Controlled vocabularies

Controlled vocabularies are also important for metadata and flexible metadata models because they can facilitate the data description process and reduce errors, ambiguities, and misunderstandings [95, 2]. In many cases, controlled vocabularies define the admissible content for a metadata element (i.e. values of the metadata element) and can be easily incorporated into automation procedures, contributing to quality control by providing users with a list of allowable entries for specific metadata elements [20].

Controlled vocabularies are an information tool that contains standardized words, terms, and phrases that denote ideas, physical characteristics, people, places, events, subjects, and many other concepts and their values [110, 2].

Moreover, controlled vocabularies are a fundamental tool that provides valuable points of access to resources during the search and discovery process [246]. They are a restricted list of values or concepts typically used for descriptive cataloging, labeling, tagging, or indexing [114]. They are tools that help organize the information in a structured form to categorize, classify, index, and retrieve information or data [2, 108].

In general, the use of controlled vocabularies helps to overcome the following limitations:

- differences in lexicon interpretation (conceptual variations);
- differences in the use of lexical expressions (social variations);
- expansion of lexical meaning (polysemy);
- lack of knowledge about the lexicon [222].

There are several types of controlled vocabularies [114, 2, 56]. A concept list (pick-list) is often used for administrative and structural metadata elements. They are often displayed as drop-down lists or as selection boxes with buttons, presenting all of the existing options. An authority file is a controlled vocabulary that includes synonyms or variants for each concept. A taxonomy is a controlled vocabulary in which all terms belong to a single hierarchical structure and have a parent-child relationship with other terms. A thesaurus is a kind of hierarchical list of terms containing synonyms or alternative expressions for each term and possibly even antonyms. Ontologies (i.e. formal, machine-readable specification of a conceptual model in which concepts, properties, relationships, functions, constraints, and axioms are all explicitly defined), header-subject lists, glossaries, folksonomies, or even classification schemes are also types of controlled vocabularies aimed at

88 https://ddialliance.org/

structuring concepts and ideas [114, 2, 56]. Furthermore, controlled vocabularies can be open-ended, where new values and concepts can be added over time [110], and closed, where no new suggestion or terms can be introduced.

Controlled vocabularies can also be classified according to: purpose (e.g., discovery, usage, semantic, syntactic); form (e.g., flat, multilevel, relational); and functionality (e.g., taxonomy, dictionary) [20].

There are various vocabularies established in the scientific community. The most popular are:

1. ISO 639 [89] which encodes names of natural languages (e.g. en for English, pt for Portuguese);
2. ISO 3166-1-alpha-2 [90] which is used for country codes (e.g. PT-Portuguese).

Moreover, vocabularies are also used in various metadata schemas [50]. For example, the Dublin Core specifies the use of vocabularies such as ISO 639, Internet Media Types [91] and Thesaurus of Geographic Names [92] for Language, Format, Coverage respectively; the Data Documentation Initiative uses a specific tool list [93]; and the Data Cite uses a list of contributors [94]. In general, a lot of vocabularies used in various schemes can be found in the Linked Open Vocabularies [95] catalog, where the vocabularies are classified according to domain and categories.

In library and information science, engineering, computer science, medicine, and health science, controlled vocabularies have already proved their usefulness by standardizing terms, increasing interoperability, and reducing errors in manual and automatic descriptions [2, 253, 130]. Controlled vocabularies also improve information storage and web navigation systems [110]. In RDM, controlled vocabularies also improve the quality of data description by improving data findability, accessibility, interpretation, and reuse [278, 134]. In other words, controlled vocabularies improve the efficiency of RDM and make the data FAIR.

In line with the increased importance of FAIR principles, controlled vocabularies should and can be FAIR, machine-actionable, improving interoperability between systems [56].

In this context, the CODATA [96], together with the DDI Alliance group [97], has published ten simple rules for turning a vocabulary FAIR in 2019. The main points among which are issues related to licenses allowing reuse of vocabularies, management mechanism, completeness, identifiers for vocabularies and domains, schema design, rich metadata, and registration of the vocabularies [56].

The more the use of vocabularies in information systems, the more benefits they bring to the scientific community. They can be used in data repositories, enhancing visibility and interoperability and facilitating data description and deposit processes [108]. They can and have been implemented in some data and institutional repositories, reducing input time and improving metadata quality (e.g., IDEALS [98]). Some institutions are developing them not only for repositories but also for other RDM systems and platforms [163, 10].

89 https://www.iso.org/iso-639-language-codes.html
90 https://www.rallybel.com/pt/links_iso_code2.html
91 https://www.iana.org/assignments/media-types/media-types.xhtml
92 https://www.getty.edu/research/tools/vocabularies/tgn/index.html
93 https://ddialliance.org/Specification/DDI-CV/TypeOfInstrument_1.1.html
94 https://tinyurl.com/ye27vvnk
95 https://lov.linkeddata.es/dataset/lov
96 https://codata.org/
97 https://ddialliance.org/
98 https://www.ideals.illinois.edu/units

2.6 RESEARCH DATA MANAGEMENT REPOSITORIES AND PLATFORMS

Since the importance of research data, their openness, publication, sharing, and reuse have become increasingly important in the scientific community and society, different types of repositories and platforms have been developed and proposed for use [190]. A data repository is the nexus between Open Science, Open Data policies, and society, where researchers can store, preserve, publish, search, manage, and disseminate their data; without repositories, data cannot be shared and reused [50].

Some of the repositories could be incorporated into RDM workflows and used only for institutional data publication; others could be generic or domain-specific repositories where anybody could deposit their data. Some repositories could have data steward support for depositing and describing data, others could use a metadata standard or flexible metadata models that researchers could use without any support. In general, all data platforms and repositories have the same goal - to make data available for reuse and increase the value of the data by citation [50, 104]. Most of the generic and domain-specific repositories can be found in the Registry of Research Data Repositories (re3data) [99] and in the Directory of Open Access Repositories (OpenDOAR) [100].

2.6.1 Dendro platform

One of the RDM platforms, incorporating flexible metadata models and controlled vocabularies, is the Dendro platform [101], development of which started at the University of Porto in 2014. Dendro is a collaborative institutional platform and open source solution for describing multi-domain datasets. Initially, its main goal was to prepare and describe data in a more detailed and qualitative way, to organize data for further publication and transfer them for long-term preservation in one of the supported repositories such as Zenodo [133, 38]. Later, Dendro implemented more features related to different phases of the data-lifecycle, such as export of datasets, DOI attribution, dataset search, among others [51].

One of Dendros' focuses has been to improve metadata quality and granularity of data description, using a flexible metadata model approach, where it is possible to combine different descriptors from different metadata standards such as Dublin Core [237, 238]. The creation and implementation of controlled vocabularies have been one of our contributions to reducing description errors, thereby improving metadata quality and simplifying the data description process [133, 134]. In particular, descriptors such as Additive, Reactor Type, Hydrolysis, Catalyst, and Reagent have been created for the Hydrogen Production domain, for which drop-down lists of terms and concepts have been added to standardize data description in projects of this scientific domain (Figure 3). A series of experiments with researchers showed that the platform was easy to use, and the approach of implementing flexible metadata models and controlled vocabularies helped to facilitate data description and improve metadata quality.

2.6.2 Zenodo repository

Zenodo [102] is a generic multidisciplinary repository where any researchers from different scientific domains can deposit and describe their projects' outputs [87]. It is an open-source repository, developed through a collaboration between the OpenAIRE initiative and CERN and launched in 2013. The main goal is to make science

99 https://www.re3data.org/
100 https://v2.sherpa.ac.uk/opendoar/
101 https://github.com/feup-infolab/dendro
102 https://zenodo.org/

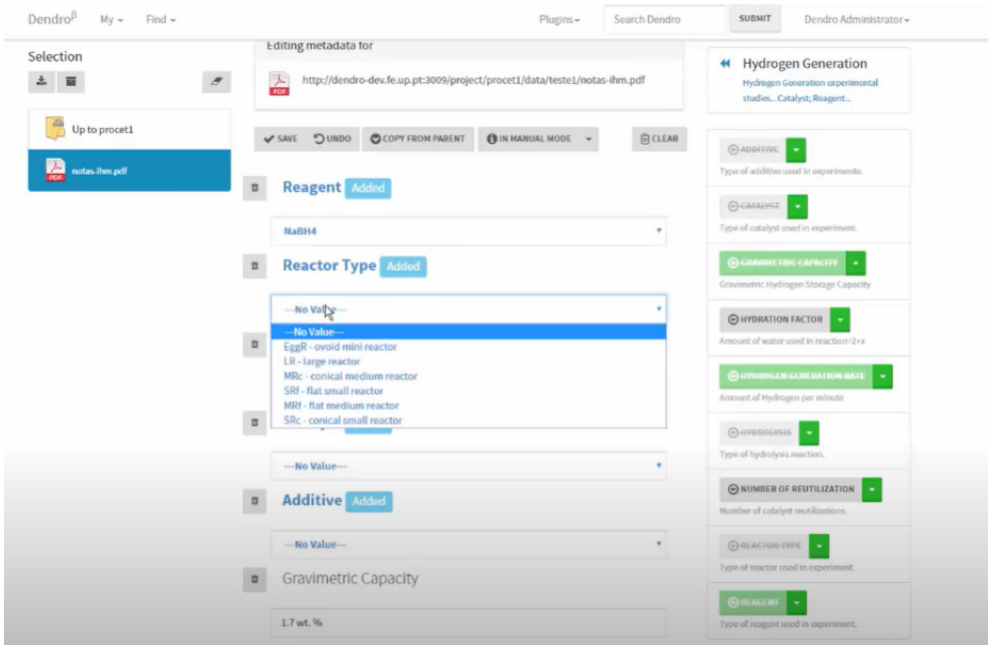

Figure 3: Data description on Dendro using controlled vocabularies

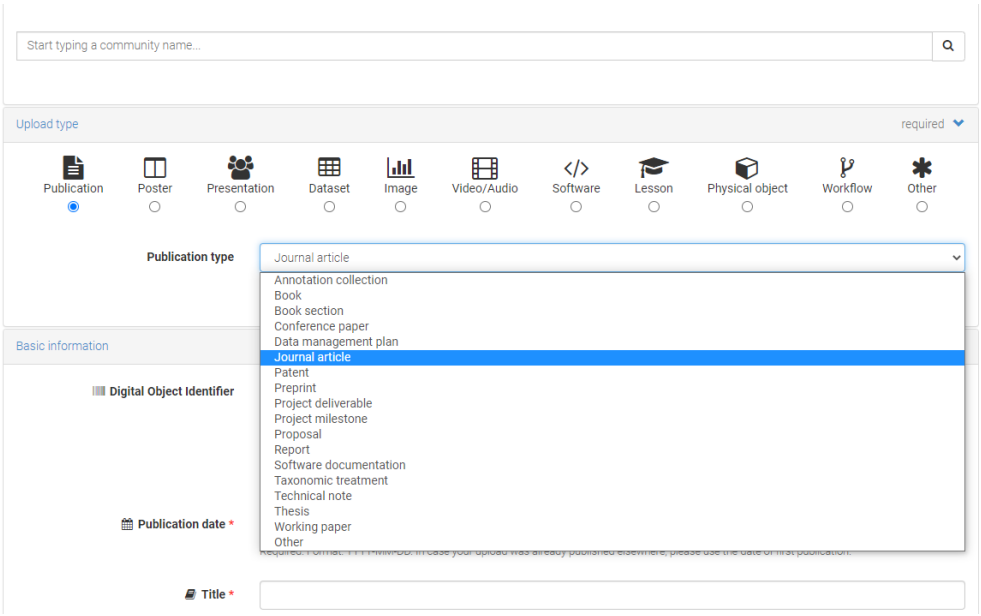

Figure 4: Data description on Zenodo using controlled vocabularies

open and provide a unique place for researchers to share their publications, DMPs, datasets, presentations, and other digital resources. In addition, various communities can be organized on it and all project participants can be brought together.

Zenodo allows the automatically DOI attribution for each digital deposited resource, making data citable and reusable by others. The repository uses the DataCite metadata schema [103] for description, which includes a small set of mandatory descriptors that can be easily used by anyone, even without specific RDM knowledge or experience. The repository also implements controlled vocabularies (e.g., for descriptors Publication Type, License, Access Rights and Grants, among others) (Figure 4). Zenodo is currently the most widely used repository among researchers with long-tail data projects.

103 https://datacite.org/

2.6.3 B2SHARE repository

The B2SHARE platform [104] was developed by EUDAT Collaborative Data Infrastructure [105] and launched in 2014. It is an open-source repository designed to provide data storage and sharing services for the scientific community with a user-friendly interface. It was been designed with users' needs in mind, with customization options. It is a generic repository suitable for research data from different scientific domains, commonly used for long-tail data [7].

B2SHARE assigns a DOI to each deposited dataset, has a small metadata set based on the Open Archives Initiative Protocol for Metadata Harvesting, and uses metadata such as author, data publication, description, and keywords. Moreover, the B2SHARE service focuses on interoperability and connectivity between different systems. That is why EUDAT has developed various APIs so that the B2SHARE service is able to integrate into various RDM workflows.

Besides B2SHARE, EUDAT has also developed other services related to data sharing and reuse. Thus, B2FIND has been created as a catalog for searching and retrieving metadata [106]; B2DROP - for data storage [107]; B2SAFE, a service that helps the community to implement data management policies [108]; B2STAGE - for easy data transfer between EUDAT resources and external computational facilities [109] and B2NOTE - for data annotation [110]. The B2NOTE service helps to enrich the metadata of datasets deposited on B2SHARE and uses a flexible metadata model approach and controlled vocabularies [151, 152].

2.6.4 INESC TEC RDM repository

Data sharing and storage are important services that can greatly benefit researchers by improving the research data FAIRness. In this context, institutions have also started to develop repositories to provide their researchers with the necessary RDM services and tools.

Thus, for deposit, sharing, and reusing research data at the Institute for Systems and Computer Engineering Technology and Science, the INESC TEC RDM repository [111] was developed under the TAIL project [112]. The data repository is an instance of CKAN (Comprehensive Knowledge Archive Network) [113], an open-source data platform built as a data management system, popular with open government data around the world (e.g., UK government and European Commission), and widely supported by the developer community [278, 43, 5, 135].

The INESC TEC repository based on CKAN provides an intuitive interface and visualization tools that make data easily accessible. In addition, it has a flexible architecture that allows customization of its features [136].

Research at INESC TEC covers many scientific domains, and the work with researchers to capture metadata requirements and design metadata models is ongoing. So, to facilitate data deposit and description processes, the use of the Dublin Core has been proposed as a domain-neutral metadata schema. However, metadata fields can be customized using key-value pairs so that we can define new fields according to researchers' needs.

As the data description and deposit processes in the INESC TEC are supported by the data steward, high-quality metadata and compliance with RDM and FAIR requirements are ensured. After "approving" the publication of the data, the data

104 https://b2share.eudat.eu/
105 https://www.eudat.eu/
106 https://b2find.eudat.eu/
107 https://www.eudat.eu/catalogue/b2drop
108 https://www.eudat.eu/b2safe
109 https://www.eudat.eu/b2stage
110 https://b2note.eudat.eu/
111 https://rdm.inesctec.pt/
112 https://www.inesctec.pt/en/projects/tail
113 https://ckan.org/about

steward deposits the dataset in the repository, assigns it a DOI and provides a well-generated citation [136].

The INESC TEC is paying more attention to RDM and funders' requirements and developing adequate and trustworthy tools and systems to support researchers in all RDM activities.

2.6.5 Data repository certification

The development and implementation of repositories are undoubtedly important for data-driven science and its compliance with new requirements that ensure openness, sharing and reuse of data. However, it is even more critical to provide trustworthy repositories that can guarantee their quality and long-term preservation. Reliable repositories will not only give researchers greater confidence in their data but will also provide greater value to the organization or institution in the research community [157, 85, 41].

In this context, the Data Seal of Approval (DSA) [114] merged with the International Council for Science World Data System (WDS) [115] in 2017 under the umbrella of the RDA to create the CoreTrustSeal Trustworthy Data Repository certification organization [116], which turned into an internationally recognized standard [68, 247, 265].

The CoreTrustSeal is an international organization dedicated to improving data infrastructure trust by providing services related to data repository certification. They also provide certification rules, requirements, and auditing services, and collaborate with many exterior organizations and RDM initiatives, such as RDA. The CoreTrustSeal activities help to guarantee sustainability and reliability of repositories, increasing repositories' reputation and making them in compliance with RDM and funder requirements. The fundamental requirements are related to the FAIRness of data with persistent identifiers and usable formats.

Due to the fact that data repositories have now been strongly encouraged to get a trustworthy certificate, CoreTrustSeal sometimes organizes calls to help institutions to turn their repositories into officially certified ones [85, 41]. In addition, they also provide a list of already certified repositories on their website [117] to help with their selection.

Regarding the certification of the INESC TEC research data repository, it is not certified yet. However, the institution recognizes the importance and necessity of this act and makes every effort to include it in the list of trustworthy ones. Along with CoreTrustSeal calls, two of our applications, "Support towards achieving CoreTrustSeal certification" and "Support towards increasing repository interoperability" have been submitted for review in order to obtain the necessary support regarding the improvement of the existing repository and its certification. Although these applications have not been approved and the repository is still not certified, work on its improvement is ongoing and will be completed in the near future.

2.7 ESTABLISHMENT OF THE RDM WORKFLOW AT INESC TEC

To support researchers in RDM activities and help to make their projects in compliance with existing requirements, along with the development of the different tools and services, institutions have started to pay more attention to RDM workflows in general.

114 https://www.datasealofapproval.org
115 https://www.icsu-wds.org/
116 https://www.coretrustseal.org/
117 https://www.coretrustseal.org/why-certification/certified-repositories/

Figure 5: RDM workflow with Dendro and B2SHARE

An RDM workflow is a "*sequence of repeatable processes (steps) through which research data passes during their lifecycle, including the steps involved in its creation, curation, preservation and possible disposal*" [1]. To improve the value of data in the long term, researchers should systematically perform management tasks throughout the data lifecycle, meaning that, among other tasks, they should describe their data on a regular basis. However, more often than not, they find themselves without adequate RDM tools, leaving them to resort to ad-hoc RDM practices supported by any tools that they have at their disposal [273], often addressing personal and immediate needs [168].

In this context, at INESC TEC, the first RDM workflow was proposed as part of the TAIL project[118], and brought together various tools to support RDM during the research data lifecycle, with particular attention to the data description requirements from different research areas [220].

Figure 5 depicts a workflow consisting of the Dendro platform, described in this chapter in Section 2.6.1 and the EUDAT B2SHARE services, described in Section 2.6.3, which interacted through an API. This connection was part of a Data Pilot established between the TAIL team, INESC TEC and EUDAT, through the second half of 2017, to allow researchers to describe their data using generic and domain-specific vocabularies through Dendro, and to import the resulting data and metadata to B2SHARE [83].

In Dendro, description ideally occured when the data were captured (Steps 1 and 2 in Figure 5), considering that pertinent information about research data could be forgotten if not recorded right away. When the data were prepared for deposit, they were sent to a data repository that met the researchers' requirements (e.g., CKAN, Zenodo, Figshare, and EUDAT's B2SHARE). Figure 5 shows a deposit in B2SHARE (Figure 5, Step 3). After depositing, users could proceed to data annotation, that time using the B2NOTE service, using tags derived from controlled vocabularies or free-text keywords and comments (Figure 5, Step 4). Overall, that workflow comprised Dendro as the tool for the organization and description of data, while EUDAT B2SHARE was tasked with publication and sharing data. B2NOTE complemented the annotation of datasets at the post-deposit stage [137].

The first Dendro + B2SHARE + B2NOTE workflow was proposed to cover important stages of the data lifecycle, reinforcing the notion that data annotation must appear in time during the research process to create good quality metadata. Moreover, annotating the data at the end of the workflow added new pathways to the data, and encouraged the exchange of ideas among researchers using notes that are more casual.

From the researcher's perspective, that workflow was aimed at being able to explore the RDM toolkit to meet the needs of users, with a special focus on the

118 TAIL's goal is to develop and test workflows for managing multi-disciplinary data in the long tail of science

publication and description of the project data. From a data steward and institution perspective, it allowed a better understanding of the researchers' needs when performing RDM related tasks and directed us to improve our RDM workflow [136].

As one of the improvements, we designed a workflow where data are directly deposited in the data repository at INESC TEC, described in this chapter in Section 2.6.4. As was mentioned above, the data repository that we developed and configured is an instance of CKAN [278, 43, 5, 135] and was proposed at the institution to provide researchers with the possibility to share and reuse their data.

In this simplified workflow we used Dublin Core as a domain-neutral metadata schema. This was considered as a prudent entry plan for researchers lacking RDM skills. We collaborated with several researchers from different domains on RDM issues [220]. Some contacts led to data description and then deposit at INESC TEC repository. The process started with the decision of researchers to share their data or to cite data in a research paper and contact the RDM team and data steward. The RDM process proceeded with the first meeting with researchers about general RDM issues and an introduction to the INESC TEC data repository. This also served to assess the familiarity of the researchers with respect to data publication and metadata standards.

To simplify the preparation of the data, we created a dataset deposit form based on Dublin Core[119]. This form is a template for the researcher to fill in. The researcher completes the form and returns it to the data steward, who validates the metadata and completes the deposit process. This approach to metadata quality [255] is based on a human assessment performed by the data steward while verifying the form.

In this context, a set of experiences were carried out by researchers from different scientific domains, and the INESC TEC data repository was assumed as an institutional data repository that could help to involve researchers in data description, deposit, sharing, and preservation. Moreover, during that work, some questions concerning controlled vocabularies were raised. Researchers confirmed that controlled vocabularies were useful tools and could be used in specific descriptors as custom fields in the INESC TEC repository. For example, we could use the ISO 639-2: Codes for the presentation of names of languages for Language descriptor or list of the different licenses (e.g., Creative Commons Attribution-NonCommercial-NoDerivs 2.0 Generic (CC BY-NC-ND 2.0)) for descriptor License.

This work was also regarded as the starting point in RDM training activities, leading researchers to understand metadata terms and standards and familiarizing and motivating them with RDM tools. Relevant conclusions were also drawn, and improvements were proposed. The main suggestions were to gradually adjust our dataset deposit form to allow researchers to select or add more descriptors; and to define flexible metadata models that allow domain-specific requirements to be considered before depositing data, leading to richer metadata. Some of the collaborations raised the needs for DMP creation [135].

Applying suggestions for improvement and making our RDM workflow in accordance with RDM, funder, and FAIR requirements, which also occurred during the TAIL project[120], the set of the RDM tools was improved and proposed as a complete and current RDM workflow to use at INESC TEC [145, 136]. The complete and improved workflow covers important stages of the data lifecycle [135] and takes into account the needs of researchers and the requirements of INESC TEC and funders. The tools illustrated in Figure 6 support researchers in different RDM activities and include the DMPOnline tool[121] for plan creation, the LabTablet[122] tool

119 https://tinyurl.com/ybbwvq57
120 https://www.inesctec.pt/en/projects/tail
121 https://dmponline.dcc.ac.uk/
122 https://github.com/feup-infolab/labtablet

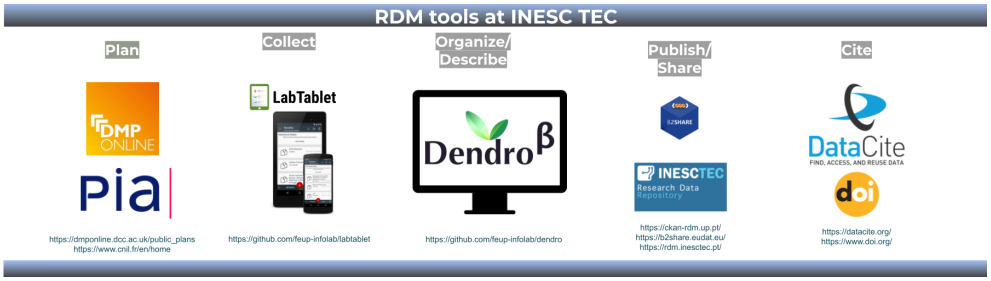

Figure 6: The set of RDM tools used at INESC TEC

for data collection, the Dendro platform [123] for data organization and description, and the research data repository of INESC TEC [124] [136, 135] for data publication.

Its further improvement is based on proposing a collaborative DMP-building method that supports researchers in DMP creation (Chapter 5), as well as on the creation of a DMP support system, which is the main goal of this work. The RDM workflow at INESC TEC is used as an example where the developing support system is applied.

2.8 SUMMARY

Open Science and Open Data movements, related policies, RDM and funders' requirements, the development of RDM infrastructures, tools, workflows and the scale of organizations and initiatives involved show the importance and efforts made towards data sharing and reuse and impact on the scientific community and society in general.

A variety of aspects, rules, and regulations described in this chapter aim to ensure that the outputs of researchers' projects comply with all existing requirements and that data are available for reuse. However, all the RDM activities that researchers must undertake are time-consuming and complicated, requiring specific knowledge and expertise in different aspects of RDM, from creating plans to choosing an appropriate repository worthy of trust. New requirements, developments, and best practices proposed by the community also should be monitored to meet the high-level of requirements for research institutes and funded projects.

In this context, institutions are making a huge effort to develop the necessary tools, systems and RDM workflows to support users in all the RDM activities that researchers have to perform. This chapter shows the importance and necessity of developing such RDM tools, services and systems to support researchers. It also demonstrates that their development is as important as the development and establishment of the RDM requirements and legislation.

We also show that all the RDM issues we study and analyze are interconnected and can be seen as a single whole, developing RDM workflows and infrastructures at institutions or improving already existing ones. The example of the establishment of the RDM workflow at INESC TEC shows how much effort institutions should put into attending all existing RDM and funder requirements, as well as researchers' needs. Moreover, this example shows that developing a well-established institutional RDM infrastructure is time-consuming, which in turn could be established more easily with a well-documented guide.

All the work we have done over the years, including the development of the Dendro platform, metadata quality improvement, creation of the controlled vocabularies, development of an institutional repository, proposal of the data deposit methods and supporting researchers in all RDM activities, have led us to the development

123 https://github.com/feup-infolab/dendro
124 https://rdm.inesctec.pt/

of the DMP support system. In other words, all our developments are the background and, in one way or another, influence the formation of a well-established RDM workflow, making any institution in compliance with Open Science and RDM requirements in general.

Nevertheless, the main aspect that led us to the development of the DMP support system, of course, is the emergence of new requirements for the submission and creation of plans, which we describe in more detail in the next chapter.

3 DATA MANAGEMENT PLAN

Ever since funders began to respond to government policies and help provide open access to scientific data, the importance of creating a data management plan has grown. Along with RDM and funders' requirements, requirements for DMP have begun to emerge, motivating researchers to submit plans as additional project documents detailing the information relevant to RDM during a project. Researchers facing difficulties in DMP creation, so, began to seek for help in creating such plans, and institutions began to develop services and tools to support them. However, as the literature shows, some researchers still do not understand what a DMP is, what it is for, when it is best created, and what it should include.

In this context, this chapter focuses on various important aspects related to DMP. Namely, we explain what a DMP is and what a good DMP should contain, we look at the importance of DMP creation and describe the requirements that exist in its respect. We also show how a DMP benefits researchers and helps make project outputs FAIR, and what additional documentation may be required depending on the data types collected during the project.

This chapter also provides information about machine-actionable DMP, the projects involved in their development, and the systems that already use them. Finally, the Domain Data Protocols are identified, which in turn could facilitate the creation of a DMP.

3.1 DATA MANAGEMENT PLAN AND ITS IMPORTANCE

DMP is a document that demonstrates a strategy of the project, proving that the received funds "yield high-quality and reusable research data" [170]. In DMPs, researchers need to provide detailed information about data organization and their long-term storage to diminish the risk of data loss and to explain how the data could be reused by others in the future. DMPs should contain information about the ways of collecting, organizing, sharing, and preserving data during and after the project. DMPs can be created for different types of data with the use of varied tools and methodologies. DMP creation is one of the ways for funders and institutions to implement "open science" principles [229, 241, 73, 186]. Moreover, the plans could be seen as a "rich source" of knowledge and be a useful tool for the development and improvement of RDM services because the DMPs analysis reveals gaps and weaknesses of institutional RDM workflows according to the funders and RDM requirements [224].

In addition, DMPs' relevance is due to the demand of the funding agencies for the inclusion of DMPs in grant applications [178, 170]. Some funding agencies require the inclusion of plans together with the project proposal, while others expect to receive DMPs within the first few months of project grants' approval [48, 45] (e.g., projects under Horizon2020 [44], National Science Foundation (NSF) [97][1], National Institutes for Health[2], et al[3,4] [66].

In particular, since 2005, the NSF started to develop DMP requirements and, since 2010, to require DMPs for all grant proposals [240]. Those plans were needed to provide information related to datasets, software, samples, and other artefacts, cre-

1 https://www.dcc.ac.uk/resources/data-management-plans/funders-requirements
2 https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-014.html
3 https://www.dcc.ac.uk/resources/data-management-plans/funders-requirements
4 https://www.dcc.ac.uk/guidance/policy/overview-funders-data-policies

ated during projects. They were a crucial part of the general projects' strategies, helping to align with funders' requirements related to reproducible and transparent science[5],[6]. The DMPs help to communicate RDM policies with the research community and to enable funders and the community stakeholders to learn about research data creation, storage, preservation and reuse practices [250, 97, 240].

Along with the NSF, the Digital Archiving Consultancy confirmed the benefits of the data planning and proposed their recommendation for DMP creation [161]. In the following years, several U.K. and U.S. funders (e.g. Bioinformatics Research Centre[7], National e-Science Centre[8], Medical Research Council[9], Office of Science and Technology Policy[10] administered DMPs requirements proposed by the NSF [204]. Some of the U.S. funders started requiring the submission of a similar document during grant applications to facilitate access and reuse of the data. The document consisted of two pages describing the principal information, such as rules, tools, and methods used for the data management [22]. The DMP demanded by the European Commission had the same requirements for Horizon 2020 projects with the principal focus on the necessity for data to be "as open as possible, as closed as necessary" and free of charge[11]. That activity affected European countries, demonstrating the importance of the DMP, turning it into a central mechanism of research data management [22, 154].

In that context, a good RDM is a vital aspect of the project, the importance of which is emphasized by funders, institutions, and governments [39, 61]. DMPs are essential for projects and play a principal role in good data management[12] [45, 78]. That is why researchers are encouraged to create DMPs, supported, and stimulated to think about RDM issues since the creation of the research proposal, improving the conducting of the project [85, 74, 73, 186].

According to the RDM and funders' requirements, a DMP should reflect on all changes occurring in the project, and be updated with detailed information, at least in the middle and final reporting. The DCC defines a DMP as an "active" that includes information related to data management throughout the whole project [127, 174, 157, 86]. The DCC also confirms that a DMP could be a useful document between project partners ensuring RDM effectiveness [61]. Besides, there are criteria of the DMP's evaluation helping to create a DMP with more details and of higher quality proposed by Science Europe and by Bishop, B. W., Ungvari, J., Gunderman, H., Moulaison-Sandy, H. [74, 85, 22]. Authors Lefebvre, A., Bakhtiari, B., and Spruit, M., along with the Science Europe proposal define three categories of the DMP's quality criteria: completeness, openness, and actualization [154]. The completeness shows that researchers have ensured all the DMP topics' presence. In other words, the DMP with little information means that researchers don't get attention to the data management issues on their project. The openness directs the data opening according to the principle "as open as possible, as closed as necessary" and FAIR principles. The actualization focuses on the constant updating and monitoring of the DMP.

Despite the fact that DMPs have a central role in RDM and help to improve RDM services, implementing funders' requirements, RDM services are not as developed as desired yet. That is why the RDM services' development, improvement, and implementation currently require a lot of attention, making institutions in accordance with all the existing RDM requirements and legislation [154].

In some scientific domains, data management is more practiced than in others. For example, researchers in such fields as astronomy, health, and genetics have had

5 https://www.nsf.gov/bfa/dias/policy/dmp.jsp
6 https://science.energy.gov/funding-opportunities/digital-data-management
7 https://birc.au.dk/
8 http://research.nesc.ac.uk/
9 https://mrc.ukri.org/
10 https://www.whitehouse.gov/ostp/
11 https://tinyurl.com/2a274vwt
12 https://eosc-portal.eu/sites/default/files/eosc_declaration.pdf

such practices for decades (e.g. preservation of the DNA [13]) [229, 154]. Furthermore, as the literature review shows, the importance of data management in projects has not appeared just now but has existed for a considerable time.

Over the past decade, data management plans have slightly changed their appearance and designation but still have the same purpose - improving the management of the projects and systems [245]. The authors Smale, N., Unsworth K., Denyer G. and Barr D. discovered the first plans similar to DMP in reports created in 1966 for aeronautical and engineering projects [13, 121]. The plans included a set of rules that helped guide the development activity.

In 1985, Ting-Peng Liang [155] demonstrated that data management could be integrated with model management to improve decision support systems and to enhance decision-making. In 1991, Clifford M. Olson [191] pointed that the plan (work plan) in a remedial investigation should include a description of logical, orderly steps (e.g. the sample rules and phased sampling) that could be adapted to computer management. Moreover, the plan could help maintain the credibility of the investigation, develop database management systems, eliminate problems associated with massive amounts of data (e.g. digitized maps), and serve as a base for the operational aspects. In 1984, Clifton L.Moss, Jr. [184] underlined the data sharing potential and importance of data availability.

Another document, similar to the current DMPs, mentioned in 1984 by Clifton L. Moss Jr., was an information system plan (ISP). It also provided information related to data management, which in turn was useful for data administration staff in preparing and executing the implementation of the tasks related to the information systems. It was seen as support for business strategy, containing details about the project, resources, tasks' schedules, descriptions, and people in charge. Data control, budget, functional specification (software, hardware), and the data information were included in the plan. In this ISP creation, the data administrator played a principal role and was focused on documentation of the data sharing [184].

In 1988, many corporations (e.g. Foothill Computer and Matrac Corporation) already attached their focus on data management, trying to learn different approaches to efficient management. They underlined the importance of well-managed data, pointing out the existence of the problems of unmanaged data in a broad range. They considered data as a valuable business resource that could contribute to business objectives. Although the data management was directed to business functions and not to scientific research, it included similar activities, such as analysis of data-related policies, processes, quality controls, standards, tools, and an indication of the data management efforts needed to be implemented [60]. The same year, in his paper "What if You don't plan?", Robert E. Wallace showed that the plan was important not only for the management information system department but for the organization in general and should have been seen as a strategic activity. He also demonstrated that inadequate planning could lead to unforeseen situations, for example, forgetting to indicate some essential aspects related to information systems in the requested budget and thus failing to make the necessary updates. That is why planning should be done with fulfilling a long-term vision and the ability to revise it as much as necessary [271]. In 1984, Stell's and Morton's work [182] underlined the importance of the storage, reusing and retrieval plan for data. They confirmed that with a good description, the data, different reports and documents could be valuable for many years, contributing to scientific development.

The plans created by the U.S. Army Toxic and Hazardous Materials Agency in 1998[14], by NASA Goddard Space Flight Center in 1993[15], and by the US Department of Energy in 1996[16] demonstrated that detailed description of the data management issues was considered as weighty. Those plans included information related to the

13 https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html
14 https://apps.dtic.mil/sti/citations/ADA460644
15 https://izw1.caltech.edu/ACE/ASC/DATA/pdf_docs/PROJECT_DATA_MANAGEMENT_PLAN_PDMP_ACE.pdf
16 https://www.osti.gov/biblio/10154215

data policy statement, the instruments, the context of the project, the data acquisition methods, the type of the collected data, the format, the volume, the data quality process, the data repositories for preservation, the standards used, the responsibilities for data management, software, hardware, etc. The plans were so detailed and elaborate that they could be good examples of well-managed data for current projects.

The importance of planning data management was also emphasized by Sanjiv Purba in 1999 [209]. In his book, he joined different authors, professors, scientists, project managers, consultants, et al., who contributed to the development of the methodology, strategy, and policies related to data management. They affirmed: *"Data management is a major driving force in the ongoing computer revolution"*. The importance of data as a valuable asset is constantly growing. It leads to the development and update of all processes, tools, methodologies, and workflows related to data management [209].

There were many different designations of DMPs; however, all listed examples are focused on data management, established rules, and activities for projects' data. As the literature shows, the importance of the data management plan has not just arisen. It appeared in the previous century and still has not diminished. It continues to grow and to spread out in different ways: scientific, research, and business projects.

Although there exist many templates, tools, and examples of DMPs, in practice, researchers continue to face difficulties in DMP creation. In most cases, difficulties arise because of the lack of specific knowledge, practice, and unfamiliarity with specific terms. A substantial number of researchers cannot provide clear descriptions about legal issues, privacy policies, technical characteristics required for storing data, and their preservation. They also cannot identify risks and possible problems that can appear during the project. In some cases, they do not realize that they collect sensitive data by conducting different types of interviews and surveys, which in their turn, usually deal with personal data [145]. That, subsequently, could require the creation of additional documents and in-depth analysis that might influence the course of the project [154]. At the same time, the researchers realize the importance of DMPs and look for support to avoid unforeseen situations. Thus, RDM services at institutions and DMP support have a huge relevance and significance [59, 179]. In addition, constant collaboration with researchers allows us to analyze different issues, motivate the researchers, improve their knowledge of RDM activities, and demonstrate the value and the benefits of the data management plan [145].

3.1.1 DMP content

A good DMP helps to manage data in their life-cycle with the long-term perspective, starting with a description of the data creation and finishing with the information about their sharing and preservation. In many cases, there are no mandatory templates for DMP creation. However, most DMP creation tools, templates and examples reveal common points [157, 86, 267, 241, 170, 231].

The analysis of the different sources related to the DMP creation and its evaluation, such as "Ten simple rules for creating a good data management plan" [170], the DMPOnline tool [17], funders DMP requirements[18] and the DMP checklist[19] developed by the DCC [71], "Request for Information on Proposed Provisions for a Draft Data Management and Sharing Policy for NIH Funded or Supported Research"[20] created by The National Institutes for Health [250], the scorecard of the DMP with different criteria of the quality of DMPs [22], and different guidance for data management plans like "Guidance document for scientists on data management, open

17 https://dmponline.dcc.ac.uk/
18 https://www.dcc.ac.uk/resources/data-management-plans/funders-requirements
19 https://www.dcc.ac.uk/DMPs/checklist
20 https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

data, and the production of Data Management Plans" [157, 86, 178], etc.[21],[22], reveals that DMPs should provide information on six principal topics: 1) project context, 2) data collection and documentation, 3) legal, ethics and security aspects, 4) storage, preservation and access, 5) roles, responsibilities, resources and costs and 6) sharing and reuse issues. All the topics are directed to the alignment of research data and projects with the RDM and funder requirements related to open science [178].

The "Practical guide for the international alignment of research data management" published by Science Europe [85], also, indicates six core requirements that DMPs should include, namely: information regarding data description, collection or reuse; documentation and data quality; storage and backup strategy; legal and ethical issues; data sharing and long-term preservation; and RDM responsibilities and resources. Horizon 2020 template for DMPs proposed by European Commission[23],[24], along with already listed topics, requires detailed information related to the FAIR data, namely: how researchers make their data findable, accessible, reusable, and interoperable [80].

Summarizing the information received from different sources and during the collaboration with researchers from different scientific domains, on which this work is based, we will highlight the ones that, we believe, cause more difficulties for researchers. They will be illustrated with some text fragments taken from the DMPs created by the researchers from the INESC TEC.

Project context is a brief description of the project, the people, entities and other project members hired or involved in the project. Moreover, in that section, the start and the end dates, information about funders, and their specific requirements related to research data management (if any) are indicated. Furthermore, it is advisable to explain why the data is valuable and whom it could be useful for [170, 127, 129, 71, 157, 178]. Here are some examples related to the project context:

"This project focuses on the study of the temporal variability of gamma radiation... The project is implemented in 2 phases: the initially planned monitoring period of 3 years, from 1 April 2015 to 31 March 2018, followed by an extension of the project duration from 1 April 2018 until 15 May 2020...The data collected and produced during the project will be useful for decision-makers, governmental organizations, regional and local authorities, national environmental agencies and ministries, as well as for other researchers. In particular, all organizations related to environmental radioactivity surveillance and air pollution control services at national and European levels"[25];

"Five responsible entities: Principal: Universidade de Trás-os-Montes e Alto Douro; Partners: Universidade Aberta; Universidade de Aveiro; Research Centre on Didactics and Technology in the Education of Trainers (CIDTFF); Institute for Systems and Computer Engineering, Technology and Science (INESC TEC)" [143].

Data, documentation, and information about the data collection. At this point, researchers should describe all types of data that will be collected, created or reused during the project, the predicted volume of data and their formats[26]. They also should indicate the methods, methodologies, instruments, and software that could be used during the data collection, as well as their processing and quality control procedures [170, 86, 71, 127, 129]. It is also important to add the specific aspects that could occur during the collection process (e.g. if some sensor left the function due to a storm) [15]. All documentation related to the data and data collection process, such as authorizations and signed agreements, should be described at this point [157, 71, 178, 127]. Some examples are:

"Data used during FARSYD include: (i) data from previous projects developed in the training (PT) and test sites (SP; DE; UK); and, (ii) data to be acquired and/or produced during the project...Several approaches will be used to obtain data, depending on the specific

21 https://www.tudelft.nl/en/library/research-data-management/r/plan/data-management-plans
22 https://www.snf.ch/en/FAiWVH4WvpKvohw9/topic/research-policies
23 https://tinyurl.com/munkmjka
24 https://tinyurl.com/2a274vwt
25 https://tinyurl.com/yx8reymc
26 https://the-turing-way.netlify.app/reproducible-research/rdm/rdm-dmp.html

target (e.g. taxonomic group; data providers and/or availability). Regarding habitat mapping, validation will be performed according to standard procedures (at least 1% of polygons to be validated). Data reflecting plant diversity will follow an approach devised with the specific objective of identifying indicator species (occurrence and abundance) in each habitat which persistence is linked to agricultural management"[27];

"The quality control procedure consists in checking if times are contiguous. If not, the missing times are inserted and the corresponding measurement flagged as NA. The time series is further inspected for the identification of potential outliers. These are in general related to the first value measured after a data gap being too low, typically due to instrumental recover after a power shortage. These values are set as missing (set as NA) in the preprocessed data"[28];

"Documentation will include: message exchanges, such as entities emails, Gmail emails, Skype and paper documents such as notes about the teams (name of participants and names of the teams), some notes in the meetings or print documents of the interviews to analyze (grids, etc). Moreover, there are PIA (Privacy Impact Assessment), Assigned informed consent, and other documents relevant to the project" [143].

Data organization. At this step, researchers should specify how they are going to organize their data and files in the dataset or datasets, whether it will be a zipped file or separate files, whether they will have only one dataset or more. In some cases, these aspects depend on the project context and the "logic" of the collection [170, 127, 129, 71, 157, 178]. To define them, researchers should decide, for example, if all datasets' files could be managed, processed, collected, reused and opened in the same way, with the same access and license or not. Here are some examples of the data organization:

"...the project will have 2 datasets: 1 - the data collected and prepared during phase 1 of the project; 2 - the data collected and prepared during the extension of the project. Each dataset will be deposited on repositories and interconnected with each other"[29];

"The data from the SAIL monitoring system are stored in the onboard computer, organised in a separate folder for each day containing the individual hourly files for each sensor..." [15].

Metadata. From the global perspective, metadata are descriptive or contextual information associated with an object or a resource [210]. It is data about data [33], that allow being discovered, reused and cited [170]. Different metadata standards help researchers with the description tasks, for example, the Dublin Core standard[30] that is the most common for the description of digital objects, or Data Documentation Initiative [31] - for data produced in the social, behavioural, economic and health sciences areas [276]. In this context, at this point, researchers decide how their data will be described, which metadata standard they will use. In some cases, a specific metadata standard is required by the funder or domain requirements, such as EML (Ecological Metadata Language)[32] e DwC (Darwin Core)[33] for the description of biological and ecological collections [276]. Some examples are:

"The data will be described according to a personalized metadata model, build on the INSPIRE metadata guidelines, assuring compatibility with the research management plan existent. To improve the fit for re-use, detailed metadata, e.g. on the temporal and spatial resolutions, spatial projection used, and the method used in the collection and analysis of the data will be described"[34];

"Metadata will be created using Dublin Core, by Dendro (specific descriptors created for specific needs) or directly during deposit stage on the data repository of INESC TEC and of the ARM Data Archive. The ARM Data Archive has specific descriptors, not metadata standards. Metadata about radioactivity sensors, or geographic location can be created automati-

27 https://tinyurl.com/2p9f5fma
28 https://tinyurl.com/yx8reymc
29 https://tinyurl.com/yx8reymc
30 https://dublincore.org/
31 https://ddialliance.org/
32 https://eml.ecoinformatics.org/
33 https://dwc.tdwg.org/
34 https://tinyurl.com/2p9f5fma

cally. Keywords are generic to domain and project, for example: atmosphere; environmental radioactivity; gamma radiation; radon"[35].

Ethics and intellectual property. In general, ethics issues depend on the type of data. Sensitive, personal or private data must be verified to the possibility of their publication. If the project includes these types of data, researchers should contact the Ethics Committee, Data Protection Officer (DPO), and/or responsible for RDM at the institution. Moreover, these types of data often require the creation of the Data Privacy Impact Assessment, Informed Consent, or other specific documents, such as additional agreements between the partners, the confidentiality, or intellectual property agreement. It is also important to identify the data author, to define how the data will be managed between all entities and project partners, establish data access, and choose the license [49]. Moreover, at this point, researchers should describe the restrictions for data, if any [145, 170], and solve the ownership issues [170, 127, 129, 71, 157, 178, 30]. Here are some examples of this stage description by researchers at their project:

"FARSYD project manages several types of data, most of which is under no specific legal requirements. Data available from previous research projects will be gathered and harmonized, and respective metadata described. In the case data is public, data and metadata will be shared with other researchers. In the case of specific datasets, namely: i) biological data collected by individual researchers within their individual projects; and, ii) data provided by the Institute of Financing Agriculture and Fisheries of Portugal (IFAP), the Integrated Administrative and Control System, IACS, and the Land Parcel Information Systems, LPIS, data is semi-public or private, respectively. Specifically in the case of IACS and LPIS, data cannot be shared even if they are anonymized. Therefore, FARSYD team cannot provide direct access to data due to legal commitments, but may provide access to metadata...Part of the data used by the FARSYD team is not a direct result of the project and thus IPR issues may apply. In the case of data produced during FARSYD, IPR ownership is from the core team (IR and task leaders), and data should be requested by the IR. Data sharing will be restricted during project development and for 6 months after project end to pursue scientific publications"[36];

"The SCReLProg Project is financed by Norte 2020 (Nº 30040), and responds to all existing requirements related to the Research Data Management and Protection of Personal Data. UTAD follows the Regulation of the UTAD Ethics Committee ("art.3º - Competências") which define a set of the rules to guide projects, including this one, which must be read and applied by all project collaborators...In case of participation in the research study, the students receive the documents with the detailed information about the project, such as the abstract, its goals, participating partners, and previous published research works. Moreover, they sign acknowledgements of understanding the finality of the treatment of their data and provide informed consent for the use of their data. This DMP has been verified by the INESC TEC DPO and is in compliance with the General Data Protection Regulation. In addition, an agreement on data processing will be created between consortium members without PIA creation" [143].

Data sharing. The way data will be shared with others is an important point that also should be described in DMP. Researchers should identify the data that could be published and preserved, decide what repository will be more adequate for their data, and define data accessibility. They can choose repository more adequate for their project, there are many different repositories: institutional (e.g. the INESC TEC repository [37]), disciplinary repositories (e.g. ICPSR for social science data [38]), or generics repositories such as B2SHARE [39], ZENODO [40] for any type of data. It is also advisable to choose trustworthy repositories or even certificated

35 https://tinyurl.com/yx8reymc
36 https://tinyurl.com/2p9f5fma
37 https://rdm.inesctec.pt/
38 https://www.icpsr.umich.edu/web/pages/ICPSR/index.html
39 https://b2share.eudat.eu/
40 https://zenodo.org/

ones by the CoreTrust Seal [41]. Moreover, at this point, researchers should define access permissions (e.g. open access without restrictions, access restricted) and/or embargo period, if any [85, 170, 178, 71]. Here are some examples from the DMPs:

"The metadata will be open on the Research Data Repository of the University of Porto (https://ckan-rdm.up.pt/). Data resulting from FARSYD will be available upon request after project publications. Metadata will be public. Part of the data (IACS and LPIS) is protected under legal commitments and will not be shared. Other data, produced within FARSYD, will be shared upon request and after consortium agreement"[42];

"The data will start to be available during the campaign. The dataset will include pre-processed data after implementation of quality-control procedures for missing values and outliers...The pre-processed data will be available unlimited. The raw data will be available upon request"[43].

Data storage, data access, backup, and data preservation. During the project, researchers collect, create or reuse many types of data. Some data are accessible by all team members, stored on the computer or external disk without any protection, and published on the research data repository, and some of them are not. That is why researchers should provide information about what data, where and how will be stored during the project or even after the project conclusion. Moreover, they should specify how the data will be backed up by identifying the schedule, rules, responsibilities for management of different backup versions, type of storage used (e.g. the external disk, institutional file-server), and permissions to access the data. As at one project, different team members may have different access to the data, it is significant to define internal data management rules. Furthermore, researchers should describe what will happen to the data after its processing or after the project conclusion (e.g. data could be preserved or destroyed). If the data are going to be preserved, then the researchers should decide who, where, and for how long will have access to them. If the data are going to be destroyed, then the researchers should choose the time (e.g. as soon as possible after the project conclusion or after the data processing). Moreover, not all data produced during the project should be and will be preserved, so specific consideration should be given to data selection, and, sometimes, it is important to explain why the data should be preserved [170, 127, 129, 71, 157, 178]. For example, if some sensitive data are going to be preserved, it is recommended to keep the data for as little time as possible. In other words, they should be preserved only for the time, strictly necessary for their processing, continuing with their destruction. Since it could be difficult to answer technical questions on DMP, collaboration with the IT department or institutional data steward is advised. [145]. Here are some examples:

"The data will be stored in a specific folder in a computer owned by the PI (Principal Investigator), with an encrypted disk using the XTS-AES128 algorithm at rest and a strong encryption key. A backup will be saved by the PI, at least every 6 months, in a protected external hard drive owned by the PI. Only the last two backup versions will be preserved and they will only be accessible to the PI. Older versions of the backup will be deleted. More details will be added in an updated version of this document" [239];

"The data will be stored and backed-up on the project-specific NAS (Networked Attached Storage). Some data can be temporarily stored in an external disk before the PI is transfer the data to the FARSYD NAS. Backups are set automatically to occur each Friday at 23:59. Data will be stored and backed up during 5 years after FARSYD completion (until 11/2018)...The RAW data collected during FARSYD will be available to download by FARSYD team members and editing limitations apply. Specifically, in the case of data acquired during FARSYD under legal restrictions to use and sharing (CAP payment data from national payment agencies), limitations to access, edit and download will apply to assure that all requirements for data use are met. In such specific cases, the PI will assure that team members analysing the data comply with all requirements. Security will be granted by:

41 https://www.coretrustseal.org/
42 https://tinyurl.com/2p9f5fma
43 https://tinyurl.com/yx8reymc

limiting the access to FARSYD NAS; and, by implementing complex passwords (12 digits, containing symbols) that will expire each 3 months"[44];

"The RAW data will be preserved the shortest time possible (the maximum three year after the project conclusion), limited by the needs to process the data. After data processing, the RAW data will be erased or preserved on the personal external disks, encrypted with a password, and with a limitation of the access for this RAW data (will be defined on the second version of DMP)"[45].

Responsibilities. There are different types of RDM responsibilities that could exist on the project. There could be some person, for example, responsible for DMP creation, for collecting or processing data, for backups, among others, or, entity, for example, the project partner is responsible for the dissemination of the data. Moreover, the responsibility should be thought out both during the project and after its completion. Therefore, it is recommended to indicate the information in as much detail as possible, right at the beginning of the project, in order to avoid unwanted unforeseen events [49, 170, 127, 129, 71, 157, 178]. There are some examples:

"The PI is responsible for the data management activity. Also, the PI is responsible for storage and backup of the data in the external disc"[46];

"Responsible for the collection of the data: Susana Barbosa (project PI) / Eduardo Alexandre Pereira da Silva (Engineering coordination)...Responsible for DMP creation: Susana Barbosa (project PI)/Yulia Karimova, (ORCID: http://orcid.org/0000-0002-1015-6709)" [15].

Resources and costs. During the project, researchers should plan resources and tools that will be required to manage their data (e.g., what is necessary to make the data FAIR) [85, 72]. The resources can be related to physical objects such as hardware or software or they can be related to knowledge, for example, some training session[47]. As the previous work showed, the majority of the researchers purchase external disks for data preservation [212]. However, some of them start to plan their acquisition towards the end of the project [145]. Along with resources, at this point, it is necessary to identify costs needed to deliver the plan, to acquire resources, and other types of charges, (e.g., payments for data deposit on repositories or storage costs), if necessary. According to the UK Data Service, there are two approaches to costing research data management activities. One of them includes all data-related activities and resources during the data lifecycle, the other one - only resources required for data preservation and sharing[48]. The Turing Way Community advises checking costs related to the storage solution, personnel, or software licenses [49], and provides a link to the checklist for costs created by the UK Data Service[49].

Additional information. Since a DMP must be a "living" document [127, 174], constantly monitored and updated according to changes in the project, it is necessary to schedule the monitoring sessions [145]. The schedule can be described as additional information and should contain an indication of upcoming plan revisions. After each monitoring session, it is important to apply all changes that have occurred in the project to DMP and name it as the following version. To ensure that each dataset will have a unique and persistent identifier, and provide access to them, for example, for verification by the funder, during the monitoring session, the researcher can add Digital Object Identifier (DOI)[50] of each dataset to the DMP as they emerge [15]) (see Figure 7).

Examples: *"The Data Management Plan will be revised periodically for update according to project changes. The first DMP follow-up happened 16.01.2019 and more information was added, namely about extension of the project. The next DMP follow-up will be in September 2019"*[51];

44 https://tinyurl.com/2p9f5fma
45 https://tinyurl.com/yx8reymc
46 https://tinyurl.com/yx8reymc
47 https://www.dcc.ac.uk/training
48 https://ukdataservice.ac.uk/learning-hub/research-data-management/plan-to-share/costing/
49 https://dam.ukdataservice.ac.uk/media/622368/costingtool.pdf
50 https://www.doi.org/index.html
51 https://tinyurl.com/yx8reymc

Dataset	Availability	DOI
SAIL raw atmospheric data (ship data + sensor data + geosensor data)	Upon request	
SAIL data - electric field (preprocessed data)	Open	
SAIL data - gamma radiation (preprocessed data)	Open	
SAIL data - visibility data (preprocessed data)	Open	
SAIL data - solar radiation data (preprocessed data)	Open	
SAIL data - oceanographic data (preprocessed data)	Open	
SAIL data - meteorological observations	Upon request	https://doi.org/10.25747/rp31-kf14
SAIL data - fish logs	Upon request	
Sagres - raw data	Upon request	
Sagres - processed data	Open	

Figure 7: List of preserved datasets with DOIs and availability rules

"According to DMP monitoring on 10.03.2020 there was the integration of collaborators in the research team, namely: a Post-Doc Brazilian researcher Mário Fontes; two students from the Universidade de Aveiro within the scope of PIC -Edu (scientific initiation) - Maria Castelhano and Eliana Curado. Moreover, The UC professor is on a sabbatical license, so a new professor was hired to replace him in this academic year (2019/2010). The new professor Pedro Pestana aware of the research project, its rules and conditions, and agreed with its realization" [143].

All of the mentioned steps facilitate the creation of the well-detailed DMP. Even if it is not possible to answer some questions at the beginning of the project, it is advisable to indicate that some aspects will be added in the future. Thus, the information will not be forgotten and can be filled in during the monitoring session of the plan [145, 146].

Moreover, there is some advice to add the DMP more quality and to make it more comprehensive. Firstly, before the DMP creation, a researcher should verify the requirements of the project funder and the RDM of their scientific domain. Secondly, they should clearly understand what types of data will be created, generated, reused, or produced during the project because it can change the project course or require some additional resources, documents, or collaborations. Thirdly, the DMP content should be structured, detailed with simple answers, and avoid imprecise statements. The abbreviations also should be explained and the data quality should be assured. Then, the information in the DMP should be accompanied by worked hyperlinks for tools or repositories, DOIs, or other information on the Internet. Different tables, images, and graphs should be used for easier understanding. Finally, the DMP should include not only information about RAW data but also about reused data [157, 86, 229, 178, 37].

To sum up, a DMP provides a context of the project, its purpose, information about the data expected in the project, their type and format, methodologies used for obtaining the data, data volume predicted, and data description standards. Moreover, it should contain information related to "long-term" availability and preservation, a strategy of data organization, access, storage, sharing, and preservation with a specification of when the data will be available, where and with what permissions or licenses, any restrictions if they exist or embargo period, ethical, legal, confidential issues, and costs related to their management. Special attention is required for sensitive, personal, private, and/or restricted data. Finally, DMP content should be frequently revised and updated if necessary.

3.1.2 FAIRness and benefits of DMP

Well-managed data help researchers to achieve good results in their project and diminish duplicating science, making their project more valuable [88, 22]. In turn, the information in a DMP and its fulfillment during the project execution help researchers to make their data well-managed, FAIR, and easier to reproduce [44]. The DMP, for its part, follows each FAIR principle [80].

Findability (F principle) is due to including detailed information about storage, and preservation of data, their DOI, description including metadata during and after the project conclusion. ***Accessibility (A principle)*** is due to detailed information about access to the data, access requirements, and identification of the communications protocol. ***Interoperability (I principle)*** is due to the definition of the metadata standards, methodologies, tools and vocabularies that facilitate interoperability between systems. ***Reusability (R principle)*** is due to the detailed description, accuracy, and reuse permissions or restrictions, which allows others to understand and use the data without additional contact with the author [146, 145, 80].

According to Smale, N., Unsworth K., Denyer G. and Barr, D., a DMP has an impact in three dimensions: professional, economic and institutional. Professional benefits help researchers make their work more visible and productive. Economic ones are related to funding, investments and data credibility. Institutional benefits, in turn, are directed to institutional policies and compliance with existing requirements, legislation, policies and ethics guidelines [245].

A good DMP helps to respond to the funding requirements during the project submission, giving the possibility of results validation. It helps to plan, prepare, and purchase the resources, tools, and equipment needed to manage data in advance, avoiding undesirable unforeseen events and predicting possible problems in the initial stages of a project. A DMP helps to ensure that all the work results will be well-managed during the project and will not be lost when the project is over [256]. Furthermore, the content of a DMP helps newly hired collaborators to understand the project information and all the project changes on their own and pass the project information without the presence of the team leader [145, 178, 37]. Overall, a DMP helps to increase research efficiency, structure project results in a better way, minimize the risk of data loss, reduce duplication of research data, and guarantee reproducibility and longevity of project results [157, 178, 37, 22, 245, 82, 199]. Moreover, DMP can help to identify weakness in RDM workflow and services, helping to improve institutional infrastructure [102].

A DMP itself can be seen as an output of the project and be published at its beginning, thus giving the project more visibility, which, in turn, can attract other researchers or partners for collaboration. A DMP can also help to understand the costs related to research data management in advance. It, in turn, helps to create a correct funding proposal, including not only the time and resources that will be needed for data collection, analysis, and publication but the time and resources for their documentation, processing, and preservation (e.g., different technological aspects related to the increasing space on the server, backup solutions, specific software with licenses) [178, 170]. The time invested in defining a good research data management strategy pays off when the time to reproduce the study results comes. Furthermore, by thinking about various aspects of data management, DMP helps to structure project development ideas better, to organize, and to prepare what is necessary for project data management in advance [178, 170, 37].

3.1.3 Time for DMP creation

A DMP is considered to be an active "living" document that can help projects' partners in research data management [61, 127, 174]. For example, according to OpenAIRE [52], the first DMP draft with a description of basic project ideas is sup-

52 https://www.openaire.eu/when-do-i-have-to-create-a-data-management-plan

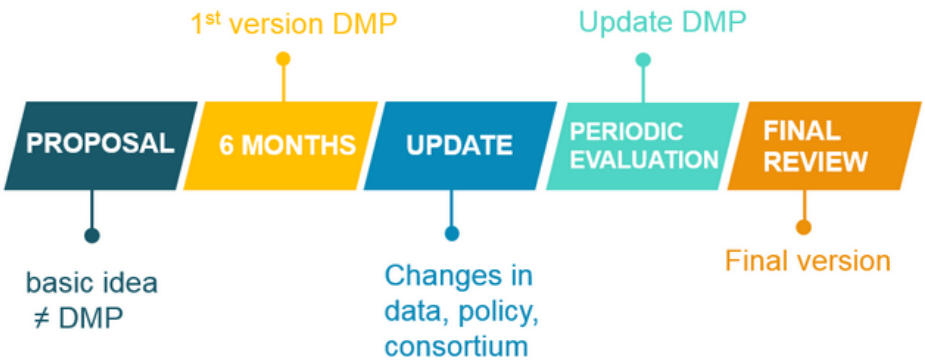

Figure 8: Timeline of DMP creation according to OpenAIRe [193]

posed to be created during proposal submission and should be improved within six months after the beginning of the project. It also should be monitored, periodically evaluated, and reviewed at the end of the project (see Figure 8) [85]. Detailing of the plan depends on the stage of the project. Thus, at the beginning stage, a DMP can have a basic minimum of information, which is gradually filled in the plan with more details throughout the project [85], because "submitting a data management plan is not the end" [154].

Projects funded by Horizon 2020 should submit the first version of their DMP (as deliverable) within the six months of the project duration, using a template (not obligatory) provided by the European Commission. The DMPs should be updated as often as possible to capture all possible changes occurring in the project. In some cases, the consortium of the project can establish a timetable for DMP monitoring (e.g., 3 in 3 months or 6 in 6 months). However, it is very important that the last DMP monitoring session should be shortly before the end of the project, document all the changes that happened during the project.

A DMP created in the early stage of the project is more beneficial for researchers as it defines the entire data management strategy of the project making data management well-organized and thoughtful [178]. Thus, the DMP becomes a useful document for all team members, partners, and collaborators [145, 256]. The sooner a DMP is created, the more useful it is [53]. That is why, in some cases, a DMP can be one of the project deliverables [170] and the project cannot be started until the DMP is approved and all the data management matters are described in detail [154, 229, 157].

In other cases, for example, according to the Swiss National Science Foundation [54], researchers can submit a DMP draft that includes an acceptable minimum of the information about data management during the project. This draft is not subject to quality evaluation criteria, however, the following versions should be improved and cover all the data management issues of the project [55]. Unfortunately, in many cases, when researchers create plans without any accompaniment or support, they do not fulfil these requirements and the DMP does not correspond to reality at the end of the project [154, 264]. Moreover, sometimes they create DMP in last minutes until the proposal submission, providing few information without request quality [102]. To sum up, a DMP can be created in different periods of the project: during the preparation of the proposal for funding; at the beginning of the project; in the middle of the project; at the end of the project; and, sometimes, during proposal evaluation for funding; or even, after project completion [145].

53 https://tinyurl.com/2a274vwt
54 https://www.snf.ch/en
55 https://www.snf.ch/en/FAiWVH4WvpKvohw9/topic/research-policies

3.1.4 Additional documents for DMP creation

There are many different projects, from different scientific domains, with different types of data, such as public or personal, sensitive and private, the latter of which can lead to the creation of additional documentation for the project [145, 80].

Personal data, according to the General Data Protection Regulation "are any information which are related to an identified or identifiable natural person", due which directly or indirectly the person can be identified[56]. It could be a telephone number, credit card, account data, etc. However, there are special categories of personal data named sensitive personal data, which require "higher level of protection"[57].

According to the Australian National Data Service (ANDS), "sensitive data are data that can be used to identify an individual, species, object, process, or location that introduces a risk of discrimination, harm, or unwanted attention. Under law and the research ethics governance of most institutions, sensitive data cannot typically be shared in this form, with few exceptions"[58]. There are many types of sensitive information, such as racial or ethnic origin, religious issues, philosophical beliefs, sexual orientation, criminal record, etc [103].

Public data, in turn, are information that can be shared and reused by anyone, without any restrictions (e.g. job descriptions)[59].

Opposite to public data, private data are not always available to the public (e.g., private data of a business company), and the same way as personal and sensitive data, private data require specific management[60].

In many cases, if the data are public, researchers do not have to worry a lot about their security. They should plan and describe all actions related to the data management in the DMP, according to the existing policies, legislation, and funding requirements, without any limitations and specific restrictions [80, 145].

More attention, of course, is demanded by personal, sensitive, and private data. For example, these data may require the creation of informed consent, with the explanation of the project objectives and data collection purposes, and the following authorization by the Ethics Committee, concerning research related to medical experimentation and human subjects [145, 96, 64]. Moreover, it can be necessary to analyze the General Data Protection Regulation [201], Data Protection Law Enforcement Directive [61], or some other national and international legislation regarding the protection of these types of data (e.g., database on data protection and privacy laws of the world)[62]) [80].

The Data Protection Impact Assessment [63], different types of agreements between partners related to the data, such the non-disclosure agreement or the confidential disclosure agreement, and other additional documents are also can be required on the project with the sensitive, personal or private data. In some cases, they can lead to a collaboration with the DPO [145].

In some cases, at the beginning of the project, the presence of such data may not be clear, and if the DMP is created at that stage and remains unchanged, there may be a lack of necessary documents and correct processing of data in the project. It can lead to unpleasant situations, even related to legislation. Thus, the DMP should be created at the beginning of the project and updated constantly [145].

56 https://gdpr-info.eu/issues/personal-data/
57 https://gdpr-info.eu/
58 https://ardc.edu.au/resources/working-with-data/sensitive-data/
59 https://www.techtarget.com/searchcio/definition/public-data
60 https://www.lawinsider.com/contracts/7NQUKaOGW0b#private-data
61 https://tinyurl.com/vpa83a2x
62 https://www.dlapiperdataprotection.com/index.html
63 https://tinyurl.com/mptc36kp

3.2 ACTIVE AND MACHINE-ACTIONABLE DMP

Normally, a DMP is a manually created text document requiring some effort for its creation [174, 172, 171, 173]. Furthermore, as it was mentioned by the Digital Curation Centre, a DMP should be "active", which in turn, requires efforts for updating and monitoring [127]. An "active" DMP can help to improve the quality of the DMP content, which is important for stakeholders involved in the RDM (funders, repository operators, legal experts, institutions, etc.) [174, 177, 172, 242]. To this end, in 2014, the RDA Active Data Management Plans Interest Group[64], [65] started to review best practices for DMP, analyze different repositories, tools, and define interfaces to identify the Requirements for Active Data Management Plans[66]. Active plans, which are sometimes called "dynamic", "machine-actionable" or "machine-readable" were proposed as "the next generation of the DMP" and were seen as more valuable for all RDM stakeholders [242, 171, 241, 173]. The "machine actionability" refers to actions of systems "to find, access, interoperate, and reuse data with none or minimal human intervention" [267, 165]. Therefore, the machine-actionable DMP helps to access DMP content, read it, interpret, exchange and act by machines [177, 172, 242].

The current DMP format (non-machine actionable) does not allow linking to different systems, therefore RDM processes such as monitoring, reporting, validation, pre-filling require more efforts, resources, and specific knowledge than if they were carried out in an automated way and using machine-actionable DMP format [173]. In other words, pre-filling DMP content obtained automatically from various systems can make this RDM task easier for researchers and other RDM stakeholders.

Based on the RDA Active Data Management Plans Interest Group effort, in 2017, the RDA DMP Common Standards Working Group was created with a specific focus on "developing common information model and specifying access mechanisms that make DMPs machine-actionable" [243]. According to the Case Statement for RDA WG DMP Common Standards, machine-actionable DMPs would help to make systems interoperable, allowing automatic exchange, integration and validation of the information [270]. Moreover, providing open, trustworthy, and reproducible science, maDMPs would benefit not only the researchers participating in the project, but also third parties, for example, funders and outside researchers who can reuse the data, infrastructure providers, researcher support, legal and ethic experts, and society in general. More specifically, maDMPs could ease DMP creation, automate suggestions of the repositories and licenses, pre-filing some fields of the plan, turning DMPs shareable, for example, with Ethical Committee or funders for validation and authorization. Machine-actionable DMP would allow plans to be readable and be acted by machines, proposing a universal format for exchanging and interoperability of the systems (e.g., preservation, processing, grants, and project management systems) [270, 243, 174].

In 2018, in this context, a first draft of the maDMP model was proposed [171]. Soon after, the RDA WG on DMP Common Standards with its more than 200 members proposed an application profile for maDMP [270]. The profile "allows information from traditional data management plans to be expressed in a machine-actionable way" [172, 242, 177]. Furthermore, there are options for "automatic exchange, integration, and validation of information provided in DMPs and facilitating the exchange of information between systems acting on behalf of stakeholders involved in the research life cycle" [176, 175, 270].

The application profile (maDMP profile) was based on the analysis of the DCC checklist [62], funder requirements, DMP themes [67], different DMP tools and requirements of the reproducible research and open science. It includes all important

64 https://www.rd-alliance.org/groups/active-data-management-plans.html
65 https://activedmps.org/
66 https://tinyurl.com/43b48uvc
67 https://tinyurl.com/2wvu246h

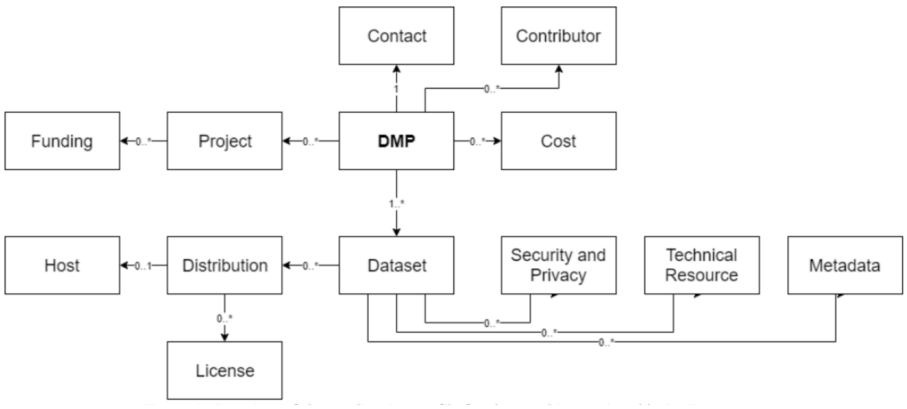

Figure 9: Concepts of the maDMP application profile proposed by RDA DMPCS [69]

aspects for creating a precisely detailed DMP that covers the full lifecycle of the project and its produced data. The maDMP profile is not a template or questionnaire. It is a *de facto* standard that helps to create a DMP in a machine-readable way, for easier reuse of the information about data that DMP includes [270] and for the exchange of information across RDM tools and systems. The profile is represented as a set of metadata elements from multiple vocabularies [176] and is an official recommendation for the research community that RDM services providers started to follow. According to RDA DMP Common Standards WG, maDMP can improve the DMP creation process, reduce researchers' efforts and turn a traditional DMP into a project evaluation tool for funding agencies, research-performing organizations or other RDM stakeholders (e.g. repository operators) [172, 177, 242]. In other words, traditional DMPs and their quality can be improved by different automation processes [175], for example, the selection of license, or opening of "closed" datasets after the embargo period. Moreover, the implementation of the maDMP can help pre-fill the DMP templates, create the first draft of the plan and increase the reuse of the information by other RDM services [73, 177, 242]. Although the maDMP profile was proposed for general domains, it can be extended with specific standards for different research domains according to their needs. In other words, maDMPs can be used in any scientific domain and be adapted to their needs [174, 173].

The application profile is flexible and includes optional fields. The minimal maDMP should include minimum mandatory fields: title, contact person, date of the creation and data modifying of maDMP, language in which maDMP is written, an indication of the existence of the sensitive, personal or ethical issues (e.g., with controlled vocabularies - yes, no, unknown) and at least description of the one dataset [177, 242]. Information about the project, dataset, its type, format, size, embargo period, timestamp of DMP versioning, DMP's state (planned or performed), license, costs, security and privacy issues, technical resources or information about funding are the other fields in this application profile (see Figure 9) [176].

MaDMPs also cover all aspects the same way as a manually created DMP. They can be seen as "automatically collected metadata about experiments". As a standard DMP, maDMP should be created at the beginning of the project and updated as many times as needed during the project. MaDMP helps to improve the reproducibility of the project results, due to easier project data management, which in turn, makes the data more FAIR. Each maDMP should have a unique identifier and be versioning. They meet all FAIR principles and can be reusable, accessible, interoperable, and reusable [174].

For standard DMP creation, the DMPOnline tool or different templates (e.g., Horizon 2020 template) are most commonly used, and usually, they are saved as PDF or DOCX documents, only readable by humans, not by a machine. Thus, implementa-

tion of the maDMP model helps to make DMPs machine-actionable and exchangeable between different systems [172].

According to Miksa, T., Simms, S., Mietchen. D., and Jones, S. the implementation of the maDMP in RDM workflows, systems and services can bring benefits to all stakeholders and reduce effort in data discovery, exchange, reuse, monitoring, evaluation, DMP creation, policy enforcement, and workflow integration [175, 173, 172, 234]:

– funders would be able to monitor compliance of the project and data with RDM requirements;

– ethics reviewers could reuse information related to the ethics issues from a DMP, so the approval process would become easier;

– similar to ethics reviewers, legal experts would be able to verify all aspects of the project related to the copyright and legal issues;

– researchers could benefit from automating the creation of DMPs and improving data management in general, for example, an easier choice of license, as it is suggested by EUDAT[68] or a recommendation of the repositories, more adequate for their data;

– connection with the datasets through DOIs in the automatic form would help publishers to control publications and their citations;

– repository operators could obtain requests and notifications for storage, costs, and other important aspects related to the deposit and sharing research data, prepare them in advance, and automate processes of the repository management;

– by exchanging information, infrastructure providers could easily access and reuse information for their goals;

– with automation of the DMP creation process, the support department workloads could decrease, increasing the quality of support and its quickness;

– project managers and institutional administrators would be able to prepare all necessary resources in advance, improving RDM services in general and reducing administrative overhead.

Bakos, A., Miksa, T., & Rauber, A., in their work, describe three steps for the creation of the maDMP. First, the data model should be defined and identify all necessary fields, standards, and ontologies that can be used. After that, during the data deposit, the form to data description appears and the user has to fill it. Finally, maDMP is generated by the system and completed with such information, as size, type, format, encoding. With automation of the DMP creation process, researchers' effort in data management and data compliance with RDM requirements can be reduced [12].

There are many groups and institutions involved in maDMP development, implementation, and testing. One case study at TUWien, for example, describes the evaluation of the system based on maDMP implementation in the institutional RDM workflow (DMap) [189]. The work shows that many of the processes related to DMP can be automated or semi-automated. However, free text and human interaction cannot be removed completely [177, 242, 172]. Moreover, connection with external services and tools and an easy interactive interface with assistance for every step turn the DMap tool into a simple and quality tool for plans created in a machine-actionable way. The results also show that researchers prefer to have prefilled parts of the DMP, use drop-down lists and controlled vocabularies, and obtain recommendations for the creation of their plans [173].

Controlled vocabularies ease the creation of the DMP, and maDMP facilitates their implementation on the systems. For example, controlled vocabulary can be a list for suggestions of the type of the data (Figure 10), estimated size (Figure 11), licenses for data reusing (Figure 12), data repositories (Figure 13), institutional names, Yes/No answers, dates, data access (open, closed, restricted). Moreover, controlled vocabulary can help to improve the quality of the information description in the DMP and diminish possible typing errors [173]. Automation of the DMP creation

68 https://eudat.eu/services/userdoc/license-selector

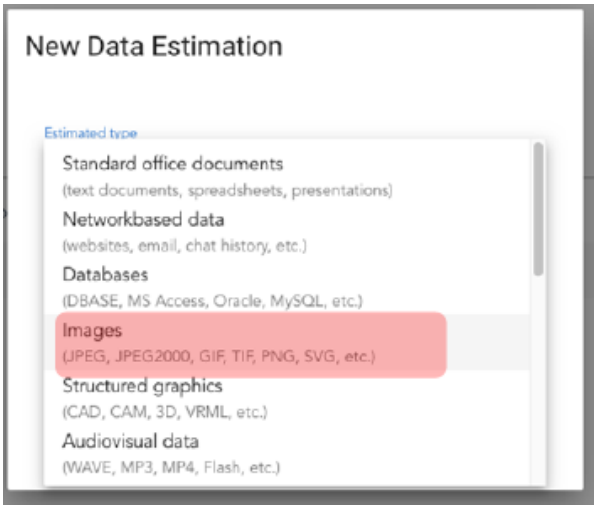

Figure 10: Controlled vocabularies for type of the data during the DMP creation [189]

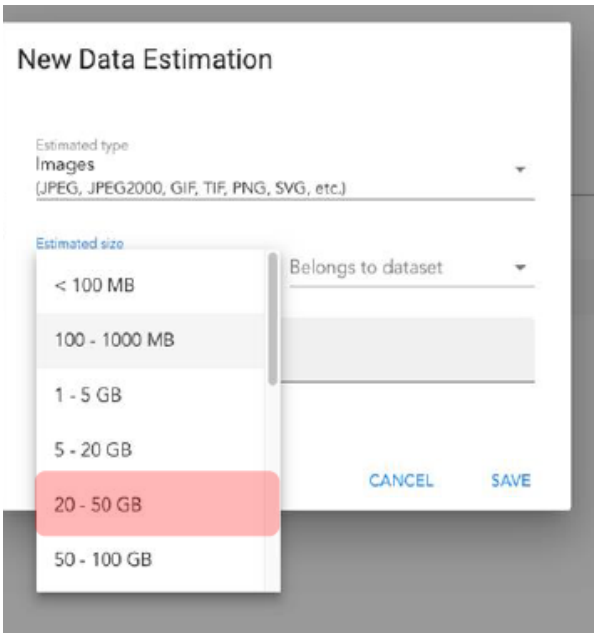

Figure 11: Controlled vocabularies for indication of the estimated size of the data [189]

is based on the controlled vocabularies and predefinition lists. It shows that controlled vocabularies have an important role in RDM activities in general and DMP creation [174].

During the definition of the maDMP application profile, a list of the activities that stakeholders have to follow, called "ten principles of maDMP", was defined [242, 177]. These principles direct to the easier application and implementation of the maDMP into RDM services and include the following recommendations [171, 175]:

1. *"Integrate DMPs with the workflows of all stakeholders in the research data ecosystem"*. This principle shows that all stakeholders should work together and look in the same direction to improve RDM services, workflows, and systems. Only in this way, the research community could receive the maximum possible benefits. Different stakeholders provide and consume information, so when information exchange is easier, its analysis is easier, and the quality is better.

2. *"Allow automated systems to act on behalf of stakeholders"*. Due to acting on behalf of stakeholders, automation can help to collect administrative data. The data obtained from other sources can also be pre-filled into corresponding fields. There

Figure 12: Controlled vocabularies as list with different licenses for data reusing [189]

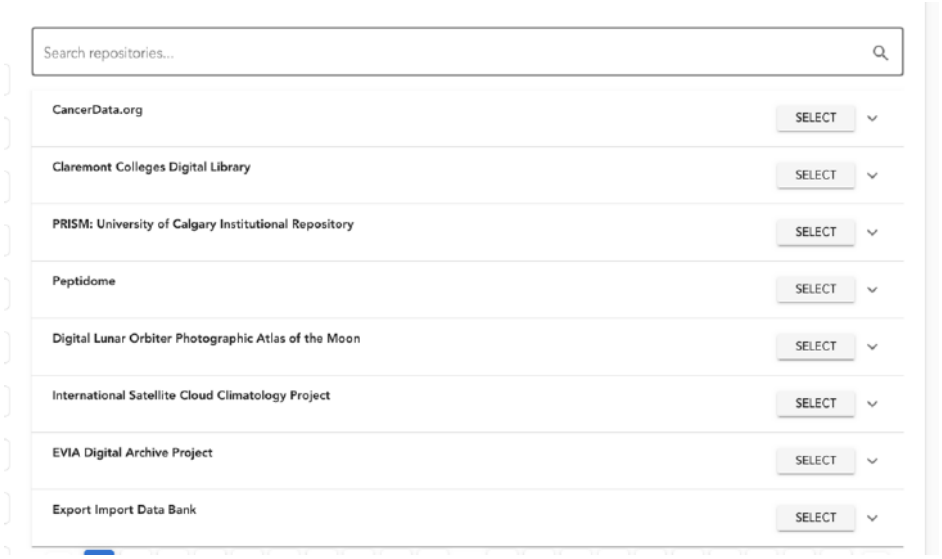

Figure 13: Controlled vocabularies as list with different data repositories [189]

are options to estimate the costs, select a more adequate license, reserve space for storage, or notificate about this necessity. After the embargo period, the data can be opened, at any time, the compliance of the project with RDM can be validated and funder requirements can be evaluated, reducing human effort and time spent for the execution of RDM tasks.

3. *"Make policies (also) for machines, not just for people"*. This principle underlines that there are many RDM policies created for researchers and institutions to comply with guaranteeing good data management and quality of the research. However, automation and implementation of the maDMP in workflows and services require the creation of the policies for machines, in a way that machines could act and execute the desirable.

4. *"Describe - for both machines and humans - the components of the data management ecosystem"*. To develop a good management ecosystem, all the necessary components (e.g., services, repositories, tools) should be well-defined, developed, and described in detail for easier use.

5. *"Use PIDs and controlled vocabularies"*. To avoid irregularity in RDM services and tools and have a worthwhile RDM ecosystem, it is advisable to use standardized terms and concepts. In this case, controlled vocabularies and PIDs can help with the organization and standardization, for example.

6. *"Follow a common data model for maDMPs"*. To have a uniform RDM ecosystem and easier information exchange, a common data model should be used and implemented by all stakeholders. This data model establishes a better connection

between systems and tools. Moreover, this model should be modular and extensible to attend to the needs of researchers from different scientific domains. As the previous point, this one also refers to using the controlled vocabularies and existing standards.

7. *"Make DMPs available for human and machine consumption"*. Due to the existence of information that can be filled in only by humans, the implementation of the maDMP should make DMPs readable, accessible, and usable by both humans and machines.

8. *"Support data management evaluation and monitoring"*. The DMP should be a "living" document that covers all aspects of the data management during the lifecycle of the project and after its conclusion. Moreover, the DMP needs to consider all changes that can occur during the project. That is why the RDM ecosystem should support monitoring and evaluation of the DMP quality, and in this case, implementation of the maDMP and automation of the processes and services helps stakeholders, for example, it helps funders to control compliance of the project with RDM policies.

9. *"Make DMPs updatable, living, versioned documents"*. DMP, in addition to being up-to-date and live, should be versioned to make it easier to keep track of all the changes that have occurred during the project's course.

10. *"Make DMPs publicly available"*. In general, DMPs can be considered as data, information about the project, and activities related to their data. To exchange and reuse this information, DMPs should be publicly available.

To sum up, workflows and systems related to RDM issues should be standardized and automated, according to RDM, DMP and maDMP requirements. The key to success is a well-balanced mode of operation between humans and machines, aiming to gain more possible benefits. Taking advantage of machines, the infrastructures will be able to improve, facilitate the effort of all stakeholders and automate processes for performing tasks related to the RDM. Moreover, automation by maDMP can help in the creation of the DMPs, their evaluation, monitoring and updating, and can add more value for the research ecosystem in general [171]. According to European Open Science Cloud deployment, nowadays, the DMP automation and making them machine-actionable has high relevance [47, 174].

3.3 PROJECTS AND TOOLS RELATED TO DMP AND MADMP

For getting all the benefits from DMPs and maDMP, the research community needs to develop RDM systems and services that support machine-actionable DMPs [177, 172, 242]. Recognizing the importance of the DMP creation, monitoring and updating, different groups, institutions and projects started to apply the maDMP application profile created by RDA WG DMP Common Standard, proposing their development after thorough testing.

3.3.1 DMPonline

Although there are several tools for creating DMPs, the main one is the DMPOnline [177, 242, 172].

In 2007, the attention of the Digital Curation Centre[69] [70] to the data management planning creation started to increase [100, 61]. In 2009, the DCC analyzed the research data management policies and verified that many UK funding agencies require DMPs during the submission of research project proposals. Such plans are supposed to describe activities related to the research data management during the

69 DCC is a *"UK service to support the Higher Education sector with Research Data Management"*
70 https://www.dcc.ac.uk/

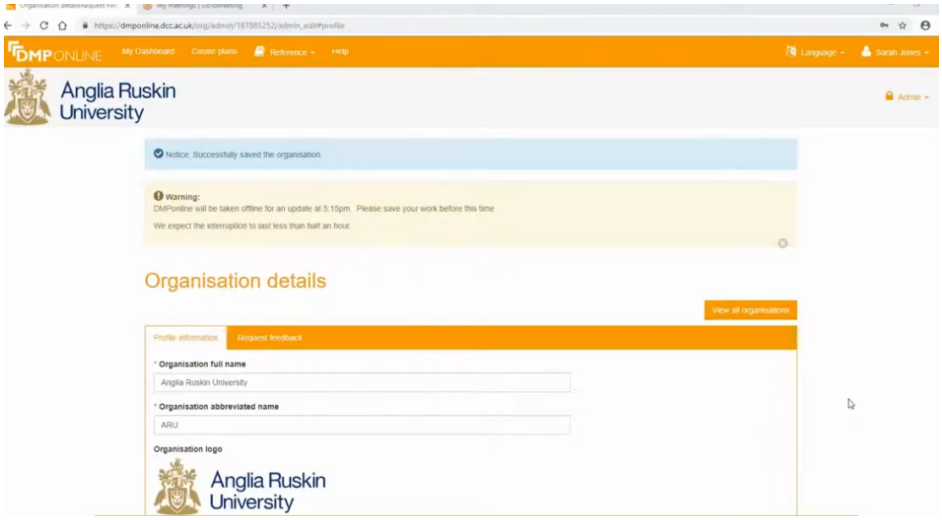

Figure 14: DMP Online interface for customization of the institutions log [63]

project to ensure that the project and the data are in compliance with the open data policies [244].

To help researchers in DMP creation, in 2009, the DCC proposed the "Checklist for a Data Management Plan" [62] based on analysis of various sources, such as the funder requirements, DMP templates, best practices in the research community, research councils (e.g., Medical Research Council, the Economic and Social Research Council) and guidance related to the research data management issues (e.g., Australian National University) [72, 128]. This Checklist consisted of the identified common themes, such as data description (e.g., type, format, and volume), standards and methodologies for data collection, ethics, and intellectual property, plans for data sharing and access, strategy for long-term preservation, among others. It helped researchers to organize data management during their project at the early stage, submit plans during the grant application stage, and update plans during the project lifecycle [72].

As a continuation of development, the first version of the DMP Online tool[71] was developed by DCC & the California Digital Library in collaboration with different librarians, researchers, and IT departments in the UK in 2010 [241]. The DMPOnline tool is a web-based platform using the DCC curation Lifecycle Model (Chapter 2, Section 2.3) and the DCC Checklist as a framework and is intended to allow researchers to improve the organization of their data during all steps of the project [70, 72, 71]. DMP Online allows maintaining and exporting researchers' plans, analyzing possible risks resulting from the data collection during the project, and planning resources in advance with an easy-to-use interface [225]. Furthermore, this tool attends to most of RDM and funders' requirements related to the DMP[72] and includes various funders' templates [85, 229] with different questions related to RDM issues and FAIR principles [177, 242, 172].

DMP Online tool[73] is used among others by many institutions of the UK, Australia, New Zealand, Belgium, Germany, France, the Netherlands, Finland, and Denmark as principal support software for DMP creation, monitoring and exporting. It is often installed as a local version in the universities, with customization of the logo and templates, according to researchers' needs (Figure 14, 15) [244].

The DMP Online tool includes various templates from different funding agencies, so the researchers can choose the most suitable one for their project and fill it [225]. Besides the templates, this tool includes guides, useful links, and detailed expla-

71 https://dmponline.dcc.ac.uk/
72 https://www.dcc.ac.uk/resources/data-management-plans/funders-requirements
73 https://www.dcc.ac.uk/dmponline

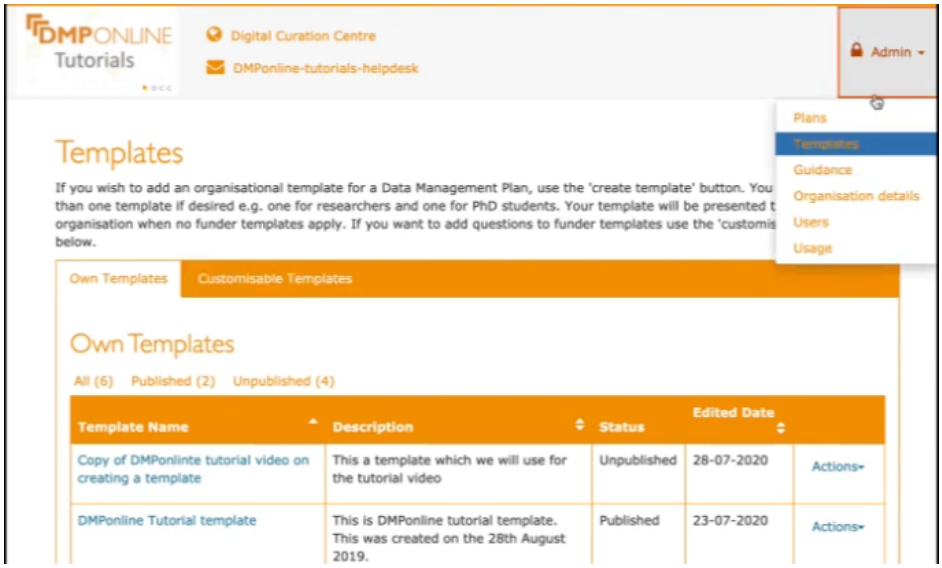

Figure 15: Customization of the template [63]

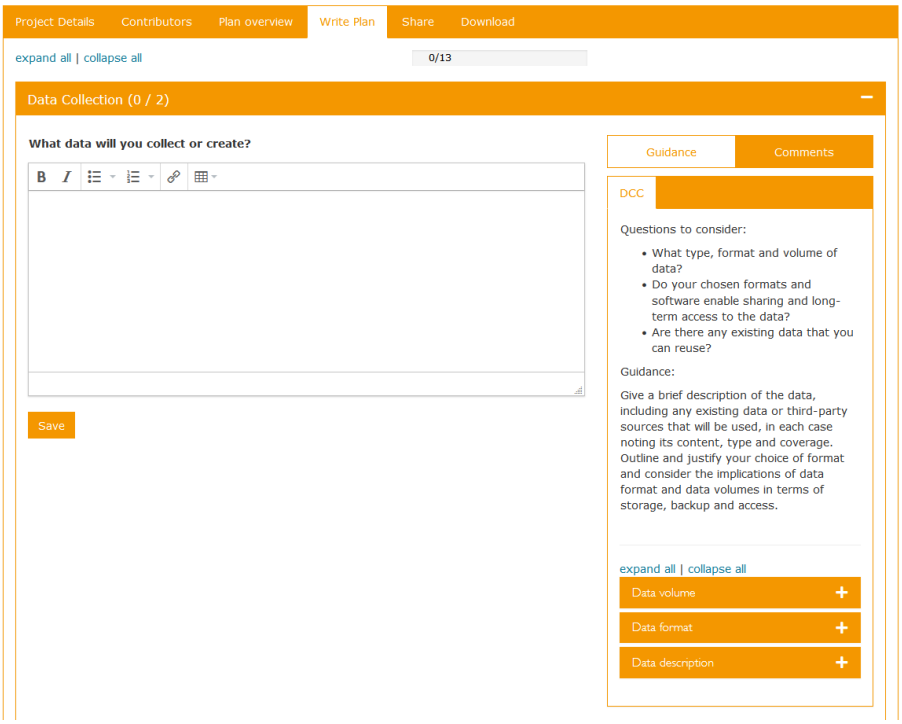

Figure 16: Detailed guidance

nations for each topic on the DMP, which helps and enables researchers to further detail their plans (Figure 16).

The DMP online tool is free to use, has a user-friendly interface, and, after a mandatory registration, a user can start to create a plan for the project by responding to all questions contained in the template (Figure 17) [225].

After filling all the forms, the plans can be downloaded, for example, as DOC or PDF documents, and further modified as desired [225].

The DMP Online tool is a very useful, universal tool, suitable for different scientific domains. In recent years, it has undergone some changes and improvements and turned into one of the principal support tools for DMP creation [244]. This tool has international recognition in different organizations and institutions. It received the award in the DPC´s Digital Preservation Awards in 2012 [244, 61].

Figure 17: Topics to be answered according to chosen template

In 2016, the DCC and the University of California Curation Center (UC3) at the California Digital Library (CDL) merged DMP Online with the DMP tool [241], resulting in a single open-source platform for DMP creation, called DMPRoadmap[74]. This platform is intended to be the base for local installation in institutions and universities, or other organizations and to be used as a support system for maDMP creation as well [243, 177, 241].

DMPRoadmap project also collaborates with groups, such as Research Data Alliance Active DMPs, DMP Common Standard, among others and is maintained in cooperation with the California Digital Library, JISC, and the University of Edinburgh. The DMPRoadmap data model complies with the maDMP standard and covers key aspects of DMP (e.g., ethical and preservation issues, metadata, and costs, for example) [244, 241]. Developers from DMP Online tool and DMP Tool see maDMPs as a key to RDM systems interoperability, which allows exchanging information contained in DMPs between different stakeholders: institutions, funders, researchers, etc [177, 241]. DMPRoadmap platform is intended to combine the positive points of both tools with the focus on the implementation of the APIs that can help to communicate with other RDM systems and repositories, making DMPs dynamic by using all benefits of the machine-actionable plans with support in different languages [244, 61].

At the moment, the DMP Online platform contains 84 795 users, 292 organizations, 93 012 plans, and 89 countries[75] and continues to be one of the principal tools for DMP creation.

3.3.2 Haplo

Another project motivated by the adoption of the maDMP is Haplo Repository[76]. This repository was developed at the University of Westminster and attended RDM requirements, allowing the management of researchers' projects in a single repository. Haplo is based on a flexible data model that can be improved according to researchers' needs[77]. Collaboration with the RDA DMP Common Standard and adoption of the maDMP application profile helps the Haplo team to improve repository operation, linking administration, project, and data management processes, enabling automation of said processes and exchanging information between all RDM mechanisms developed on Haplo. In other words, Haplo is a complete information system that links different functionalities related to the RDM issues in

74 https://github.com/DMPRoadmap/roadmap/wiki/
75 https://dmponline.dcc.ac.uk/16.02.2022
76 https://www.haplo.com/
77 https://www.haplo.com/repository

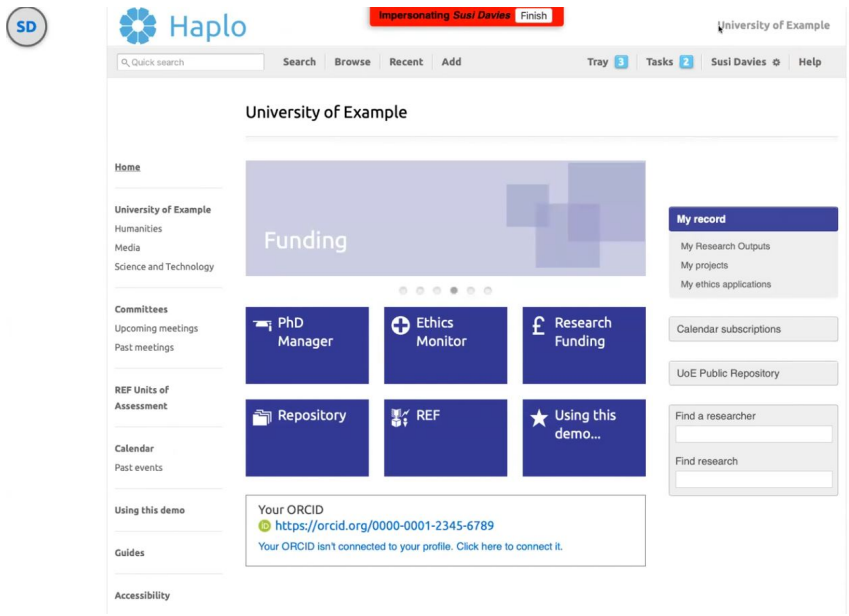

Figure 18: Interface Haplo system

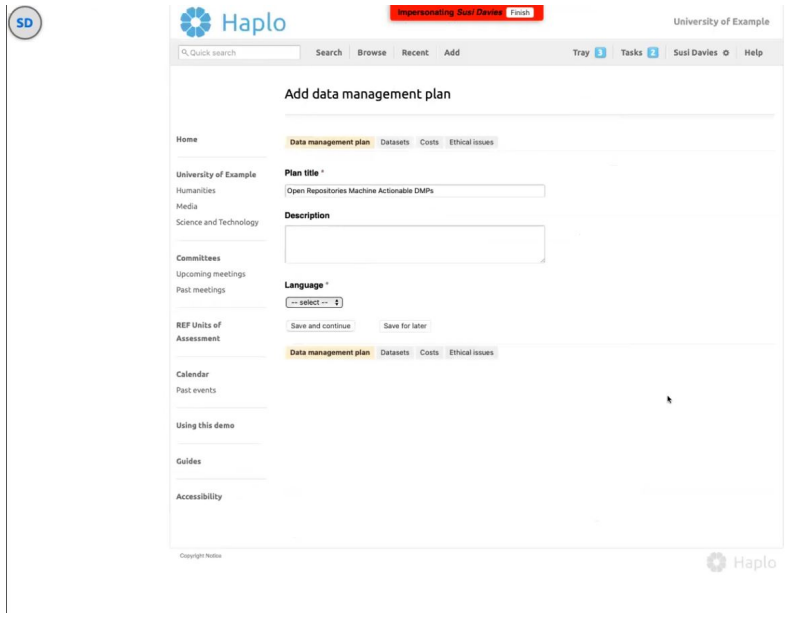

Figure 19: DMP creation process on Haplo

one place (e.g., human resources, project management, funding, DMP creation, updating, monitoring, ethics monitor, deposit data, etc.) (see Figure 18) [78].

After creating a researcher's page with personal and professional information, they can manage their projects, require ethics authorizations for projects, add, monitor, edit, and update DMP, describe, organize, and deposit their datasets (Figure 19, 20).

Haplo has impressive notification mechanisms, which allow warning or reminding researchers about any missing documents or actions during the deposition of datasets or linked projects with DMP (e.g., activating a notification about personal data and their restrictions during the deposit process) (Figure 21) [218].

Furthermore, a DMP created on Haplo can be linked semi-automatically to the dataset with the opportunity to select DMP for dataset (Figure 22) to follow and con-

78 https://www.youtube.com/watch?v=7PnFHBX3c0E

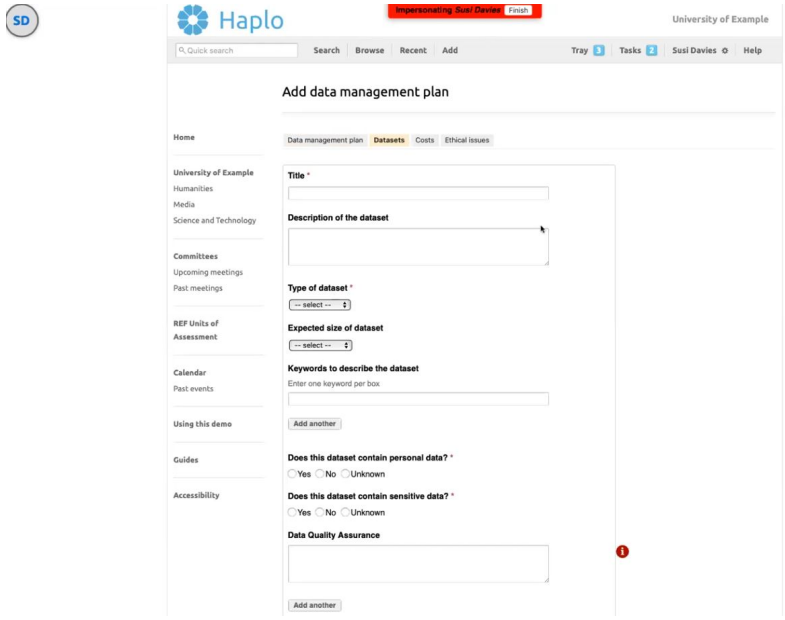

Figure 20: Dataset description on DMP

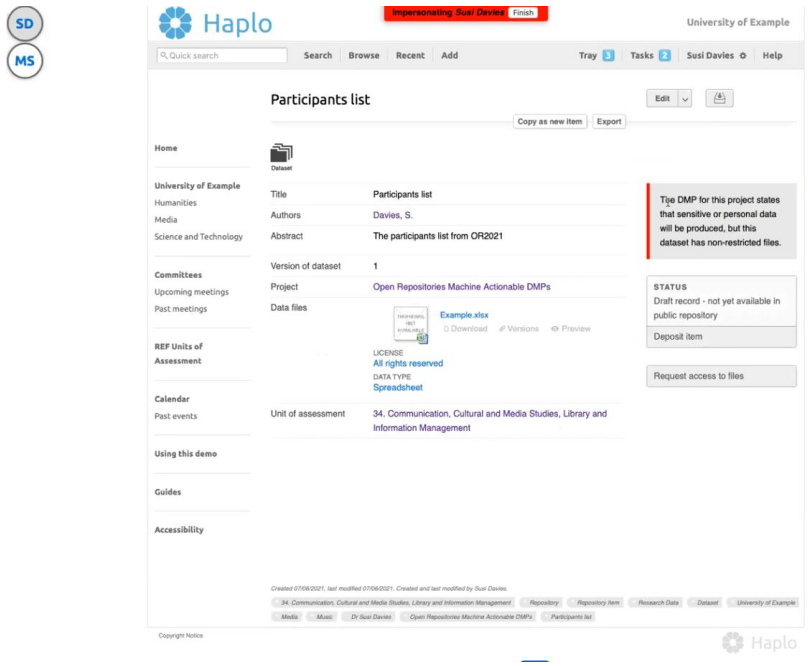

Figure 21: Notification about existence of the personal or sensitive data, indicated during the DMP creation

trol the dataset status (e.g., deposit, deposit closed access, return to the submitter, put on hold, reject) (Figure 23).

All versions of the DMP are preserved, can be opened, edited, and saved to keep track of all changes of both the DMP and the project (Figure 24).

Finally, there is the possibility to analyze all DMPs created on Haplo (Figure 25), check for errors, create reports and analyze usage statistics (Figure 26), proving very useful to institutions in general.

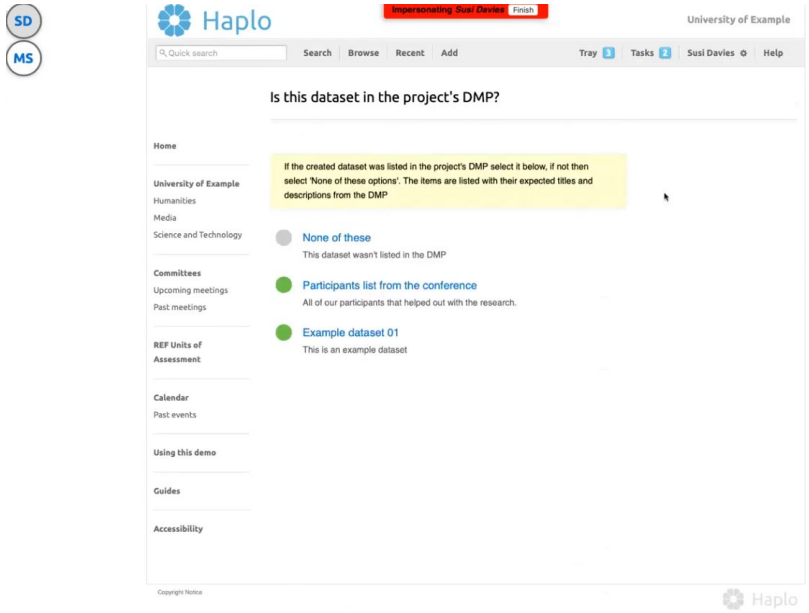

Figure 22: Connection dataset with DMP

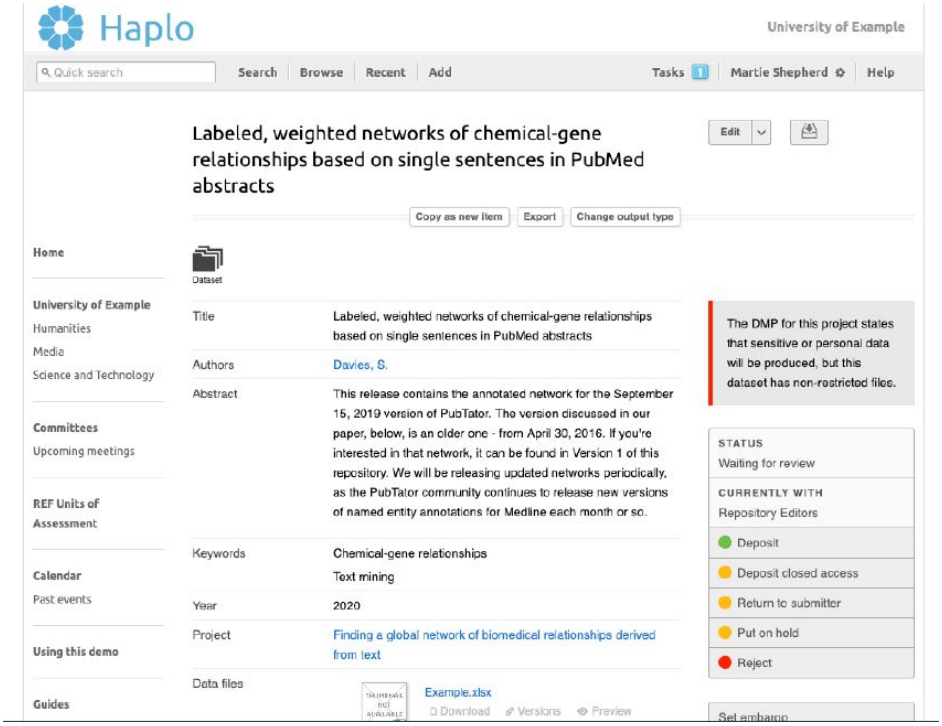

Figure 23: Dataset deposit status

3.3.3 DMP OPIDoR and maDMP4LS project

The DMP OPIDoR[79] is a platform for DMP creation, based on DMPRoadmap code[80], which unites the DMPTool and DMP Online platforms. The DMP OPIDoR is developed for medical biology researchers in the French community, together with the French Bioinformatics Institute[81] and the Institute of scientific and technical in-

79 https://dmp.opidor.fr/
80 https://www.dcc.ac.uk/news/dmproadmap
81 https://www.france-bioinformatique.fr/en/home/

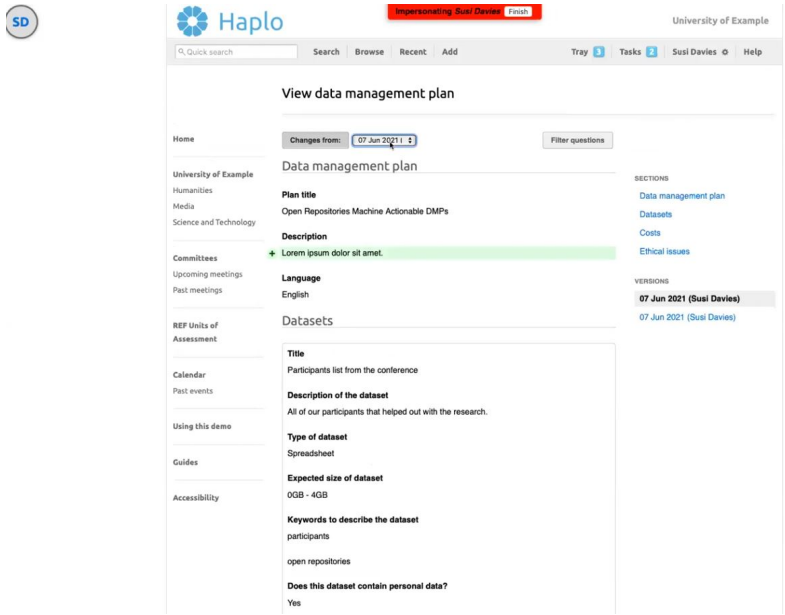

Figure 24: Changes and versions of DMP

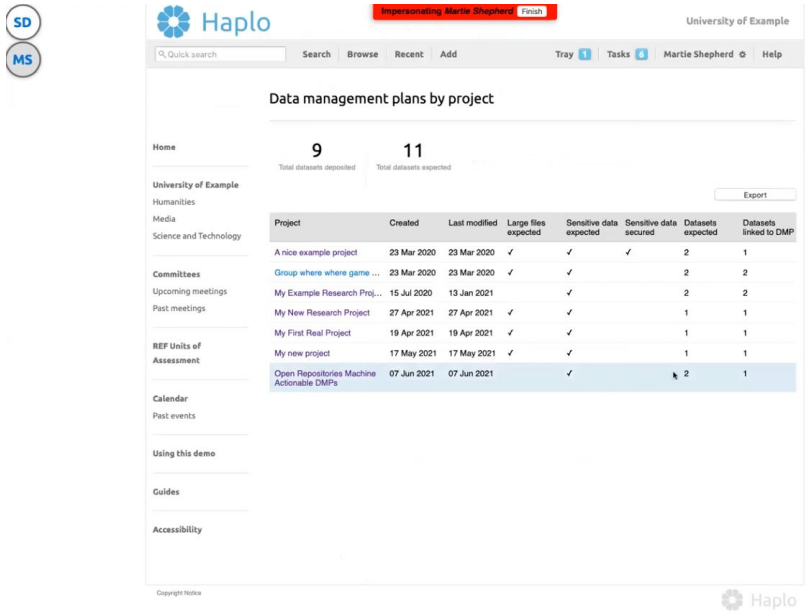

Figure 25: Report of DMPs created on Haplo

formation[82]. This platform is in accordance with best practices, RDM requirements and FAIR principles. Therefore, it guarantees researchers good management of their data. Collaboration with the RDA DMP Common Standard WG and RDA Active DMP group led to the creation of "the machine-actionable DMPs for Life Sciences" (maDMP4LS) project[83][84], which focuses on improving the existing RDM services and tools, and implementing the maDMP concept. This project is associated with the DMP OPIDoR platform[85] and will try to make DMPs readable not only by humans but by machines as well, facilitating and automating different processes during the DMP creation. It is combined with special attention to monitoring of DMPs

82 https://www.inist.fr/
83 https://anr.fr/Project-ANR-19-DATA-0017
84 https://scanr.enseignementsup-recherche.gouv.fr/project/ANR-19-DATA-0017
85 https://opidor.fr/madmp4ls/

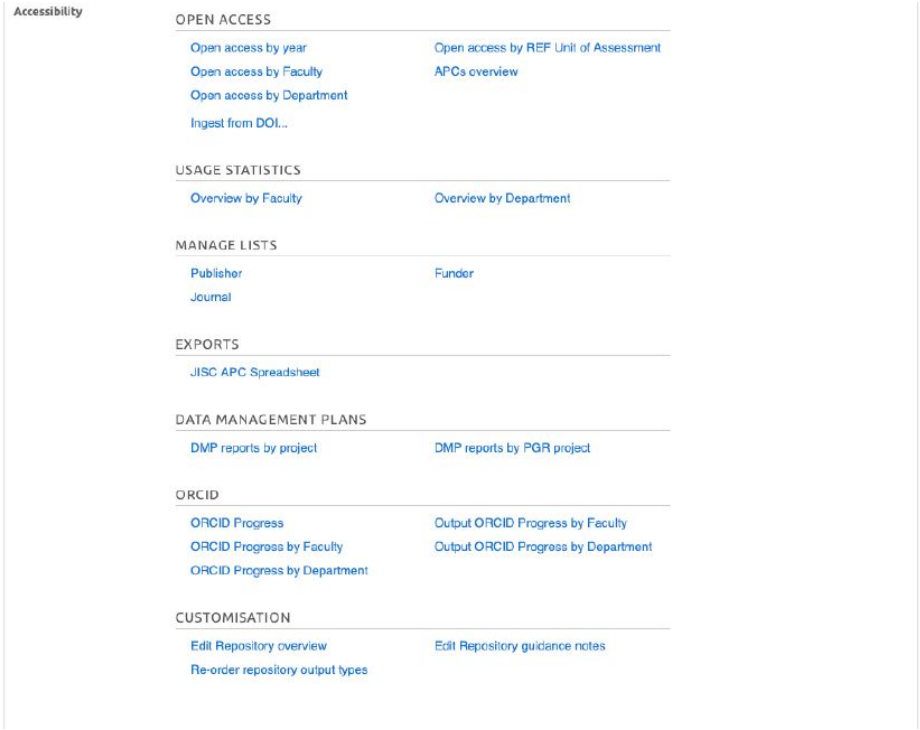

Figure 26: Statistics

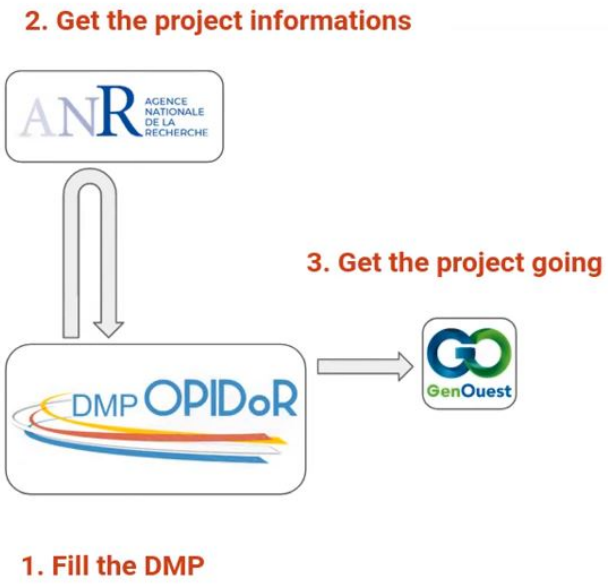

Figure 27: Steps of the RDM system in French scientific community [90]

and data storage, increasing the quality of metadata, and keeping compliance with FAIR principles [171, 90].

In general, their RDM system includes three steps: the creation of the DMP on DMP OPIDoR[86], the collection of the information about the project from Agence Nationale de la recherche (ANR)[87], and the management of the project on the GenOuest My Account Manager platform (GenOuest)[88](Figure 27) [171, 90].

86 https://dmp.opidor.fr/
87 https://anr.fr/
88 https://my.genouest.org/manager2/login

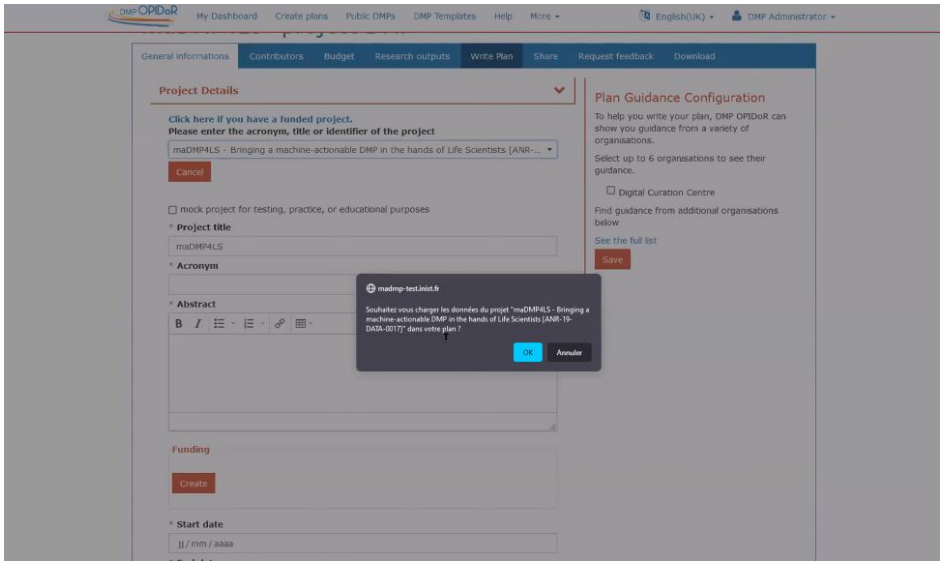

Figure 28: Connection between systems [90]

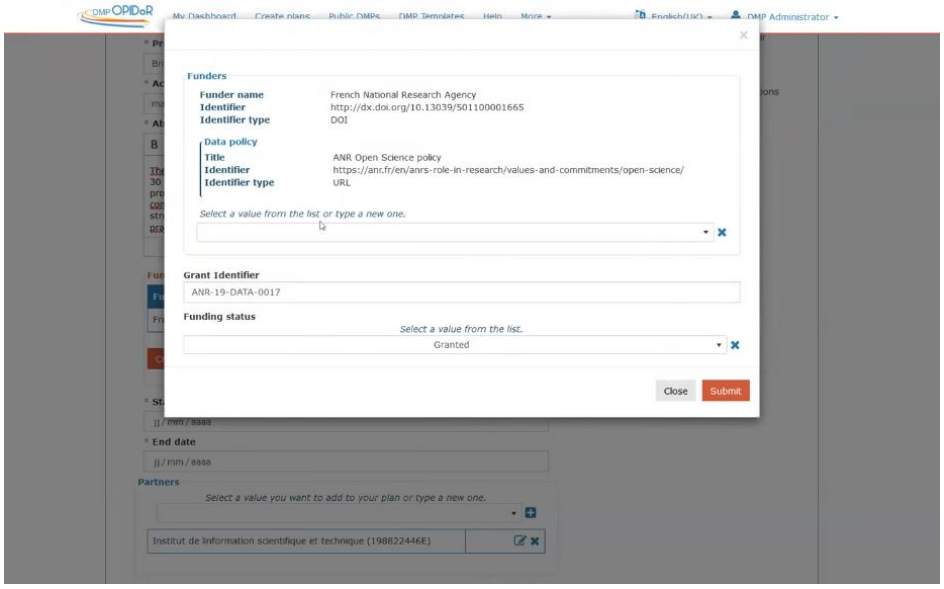

Figure 29: Information about funders is filled automatically by using maDMP concept [90]

The interface of the OPIDoR platform is similar to that of DMP Online. However, it has additional features that enable the automation of some actions. The application of the maDMP can improve the platform's functionalities and ease a possible information exchange between systems of OPIDoR, the ANR, and the GenOuest. One of the examples of this connection is related to the automatic filling of some fields, like abstract, funder, and partners, according to the description of the project that existed on the ANR (Figure 28, 29). On the platform, it is also possible to fill out the information about the Project coordinator and Contributors and assign them to different types of roles (Figure 30) [171, 90].

Using the DMPRoadmap code and existing templates as a base, the team of the OPIDoR improved some storage and backend-related aspects. Figure 31 has presented a mechanism that allows the calculation of data storage costs required for data storage in the projects. Moreover, on the same page, researchers can send warnings to the IT staff to prepare the requested size in advance. If the storage request includes the need for a backup, the automatic filling of the backup policy is also required (Figure 32) [171, 90].

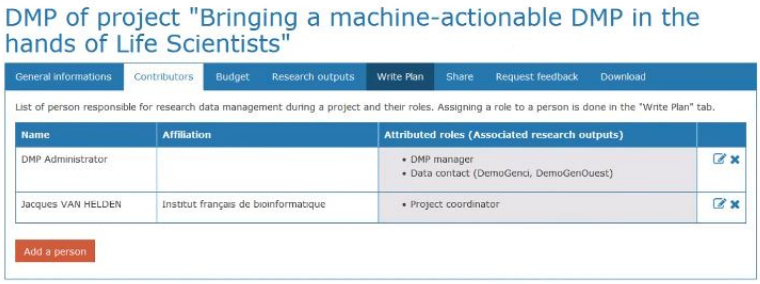

Figure 30: Information about team members of the project [90]

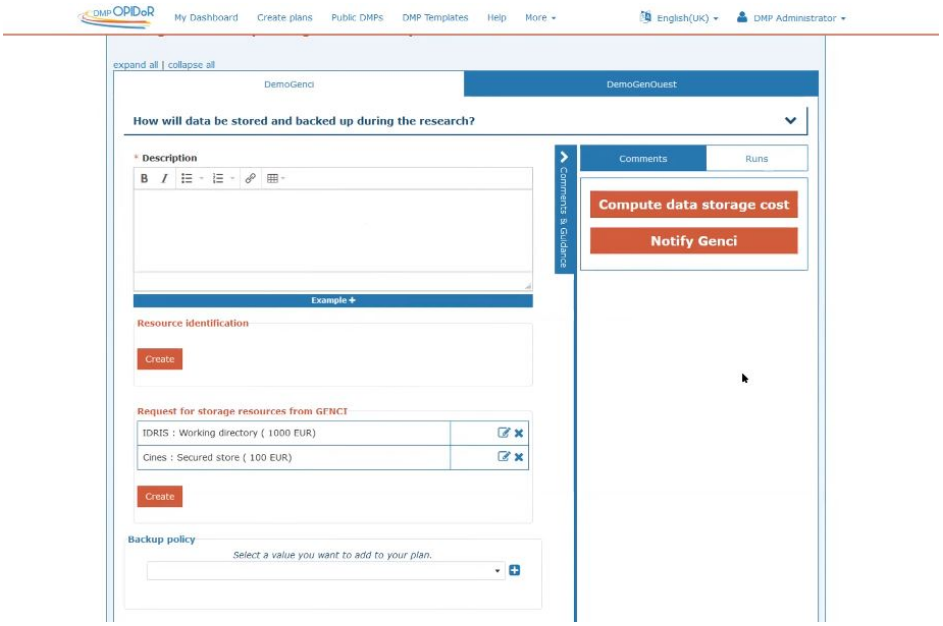

Figure 31: Interface of the calculation of the data storage cost and notification staff [90]

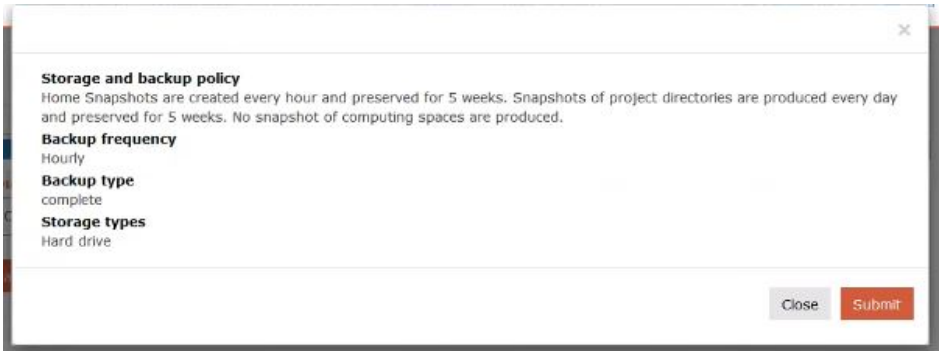

Figure 32: Automate filling of the Backup policy on the platform [90]

After successful creation of the DMP, it can be exported in JSON format and imported to the project management platform GenOuest[89] (Figure 33). Both the administrators of the platform and IT staff can analyze the newly created DMP, accept it as it is, reject it, or edit some information in it. For example, the requested size of the storage can be diminished or increased, depending on the project analysis (Figure 34) [171, 90].

89 https://my.genouest.org/manager2/login

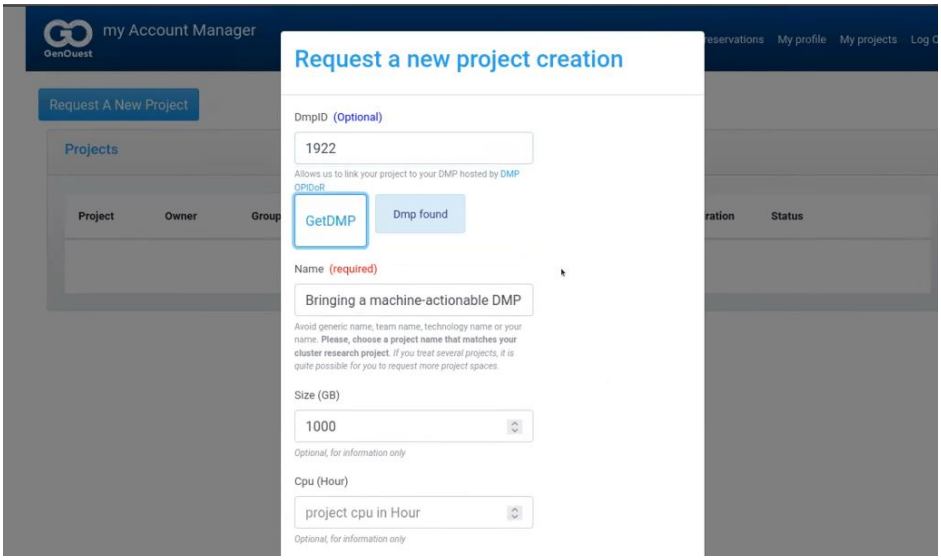

Figure 33: Import maDMP into the project management platform [90]

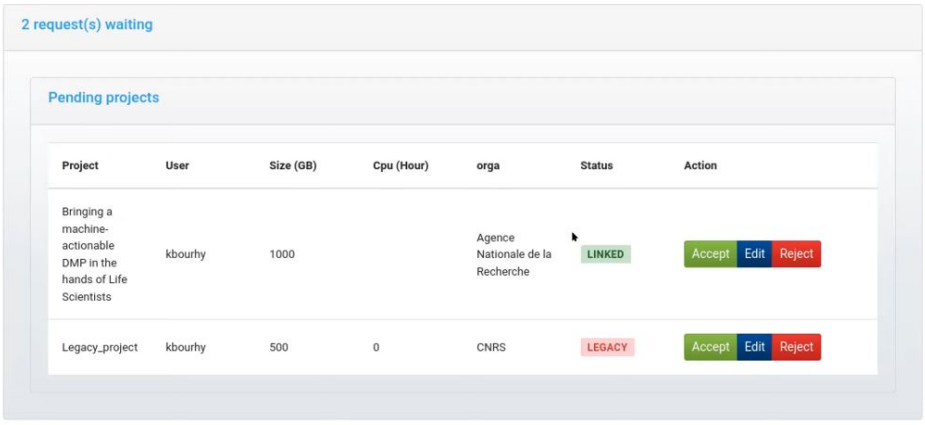

Figure 34: Interface of the DMP for accepting, edition or rejection [90]

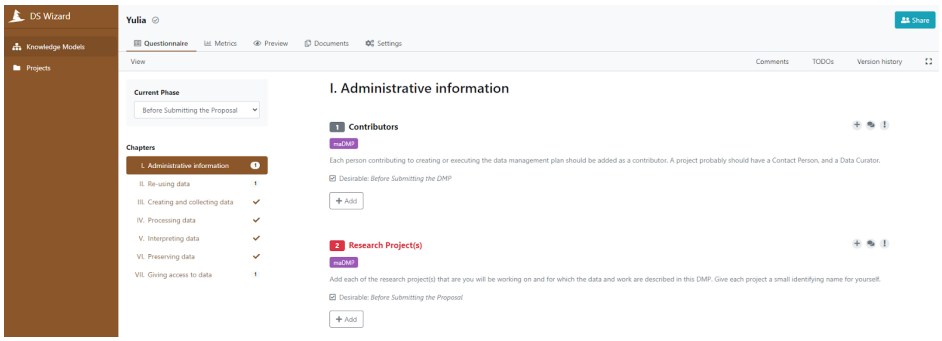

Figure 35: Data Steward Wizard interface [76]

3.3.4 Data Stewardship Wizard

Another tool that helps researchers create DMP according to the funder's requirements by choosing a more adequate to their project template is the Data Stewardship Wizard (DSW)[90] (see Figure 35). It is an online tool where researchers can create machine-actionable DMPs aiming to provide high-quality FAIR data [243, 177, 208].

90 https://ds-wizard.org/

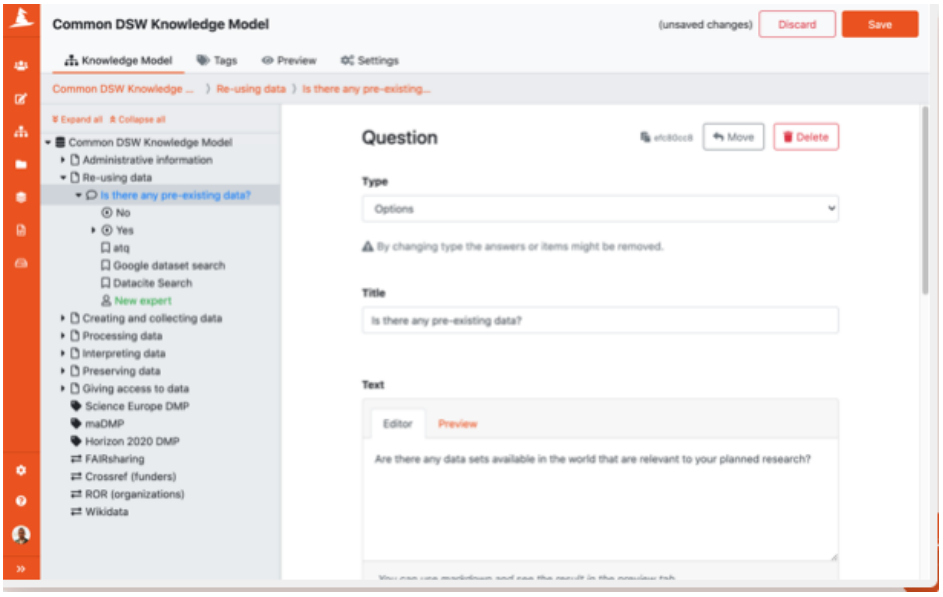

Figure 36: Interface of the DSW Knowledge Editor [76]

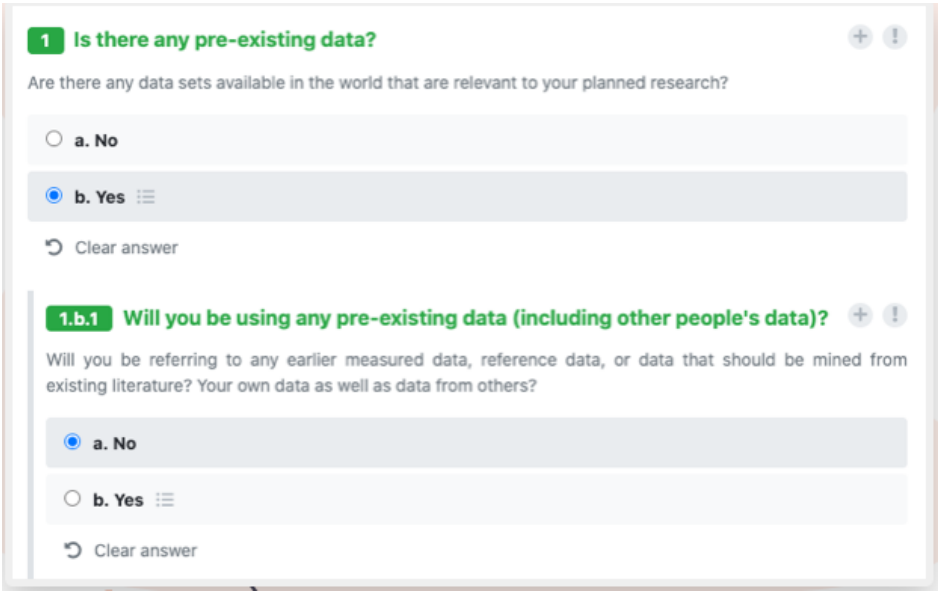

Figure 37: Smart questionnaires [76]

The DSW was developed at the Dutch Techcentre for Life Sciences together with ELIXIR CZ Czech National Infrastructure for Biological Data[91] and ELIXIR NL Dutch Techcentre for Life Sciences[92] and is based on the dynamic web forms system for DMP creation[93]. One of the specific aspects of these tools is the possibility of creating knowledge models, for example, according to the universities, or reuse of existing models, or parts of the existing ones (Figure 36) [177].

Moreover, the conditional questions are configured to help fill out the questionnaire in a "smart way" on the DSW. In other words, each next question depends on the previous answers and can be shown or not[94] (Figure 37).

Just like the other DMP tools, the DSW provides guides (e.g. SciLifeLab provides national and life-sciences specific guides) [235], hints, and links for useful resources

91 https://www.elixir-czech.cz/
92 https://www.dtls.nl/elixir-nl/
93 https://ds-wizard.org/about
94 https://ds-wizard.org/data-management-plans

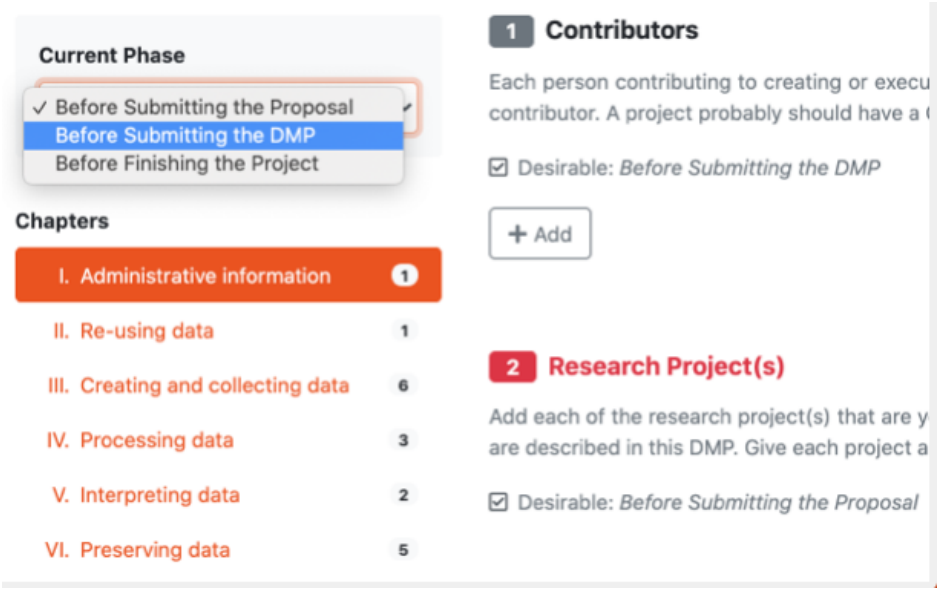

Figure 38: Project stages [235]

during the creation of the DMP [243]. Moreover, the plans created on this tool can also be shared with other researchers and saved in different versions to track changes.

Another specific aspect of this tool is related to the project stages. DMPs can be created before submitting the proposal, before submitting the data management plan, and before finishing the project. Quantity of the questions for answering changes due to the stages (Figure 38).

Like other DMP tool developer teams, the DSW team also participated in the Hackathon[95] to improve their import and export mechanisms based on maDMPs [251, 241, 177]. Moreover, they wanted to improve their root knowledge model according to FAIR principles and maDMP application profile, recognizing maDMP as a standard that helps to enrich standard DMP and to establish a way for better information exchanging[96] (Figure 39). Thus, for example, researchers can send their DMP created on the DSW to the DMPOnline tool [177].

Summarizing, the DSW is an open-source tool with focus on flexibility, openness, different guidance and user-orientation that helps to automate DMP even more with minimum effort for writing [164].

3.3.5 Argos

Argos is a collaborative, extensible, and online-machine-actionable tool developed by OpenAIRE and EUDAT that aims to help researchers in DMP creation, monitoring, and updating [177, 197, 192, 195, 81, 208]. OpenAIRE is a European project related to Open Science and technical infrastructure, providing researcher community training sessions, tools, and the connection between data providers, with a general focus on RDM aspects [97]. Moreover, OpenAIRE directs to improving reproducibility and integrity of research [196]. EUDAT, in turn, is a Collaborative Data Infrastructure related to the research data management in Europe[98], which currently proposes a suite of services to address the full lifecycle of research data.

Argo's developers see maDMP principles as the opportunity to get more benefits from DMP support tools and for RDM services in general. To adopt these princi-

95 https://www.rd-alliance.org/rda-hackathon-madmps
96 https://ds-wizard.org/machine-actionability
97 https://www.openaire.eu/
98 https://www.eudat.eu/

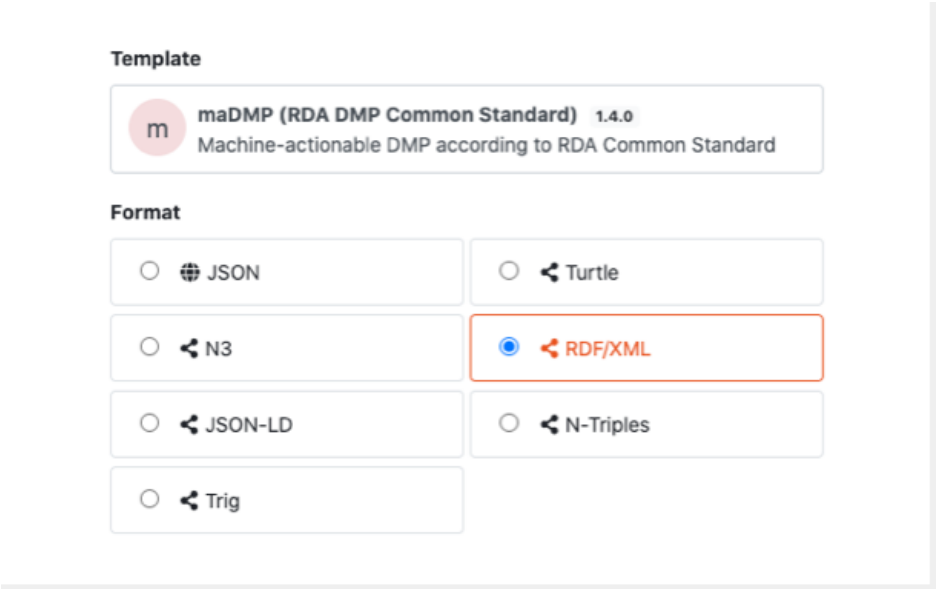

Figure 39: maDMP export in the DSW tool [76]

ples, fields of the Argos were mapped with the maDMP standard. Thus, the Argos tool is in compliance with maDMP principles proposed by the RDA DMP Common Standard, allowing information exchange between different DMP and RDM services. Furthermore, the same as most DMP tools, Argos attends RDM and funders' requirements, open science policies, and FAIR principles, and bring benefits to different stakeholders: researchers, funders, research communities or universities [198, 192, 195, 208].

Researchers, for example, can share their plans with their colleagues; engage with the RDM issues and FAIR principles, with a user-friendly interface translated into different languages (e.g., English, Portugues, Spanish, German, etc.). Using funder templates, researchers diminish their effort in the DMP creation. They can create plans at the beginning of the project, as well as during the project and at the end of the project, export, and import plans to other DMP and RDM tools. Argos is suitable for different domains and free to use [192, 195, 198, 196].

Funders, in turn, can create their own templates and make them available and mandatory for researchers. Moreover, they can customize guides and link plans with the Funder Monitoring Dashboard [192, 196, 197].

Enrichment of the templates collection (e.g., H2020, H2020 for archaeology, Science Europe, Academy of Finland, FWF - Austrian Science Fund, etc.), customization of the Open and FAIR guides for different disciplines, the configuration of the APIs, connection with Research Community Dashboards, and with other RDM services in institutions workflows, for example, with the repositories for deposit datasets, are some of the benefits that Argos can provide to the research community [198, 196].

As well as the Haplo, DMRoadmap, and Data Stewardship Wizard tool, Argos is open-source software and can be used as a standalone service. It is based on the OpenDMP and can be downloaded on the EOSC portal[99]. Argos allows creating, exporting, and importing maDMPs, making configurations to better fit researchers' needs, connecting with the EOSC services, plans publication, and using them in different RDM services [192, 195, 81, 196, 177].

The specificity of the Argos model consists of the manner in how DMPs and datasets sections are organized (Figure 40) [100]. More specifically, Argos has two main editors: one for DMP and one for datasets linked together. DMP section

99 https://marketplace.eosc-portal.eu/services/argos
100 https://argos.openaire.eu/splash/resources/user-guide.html

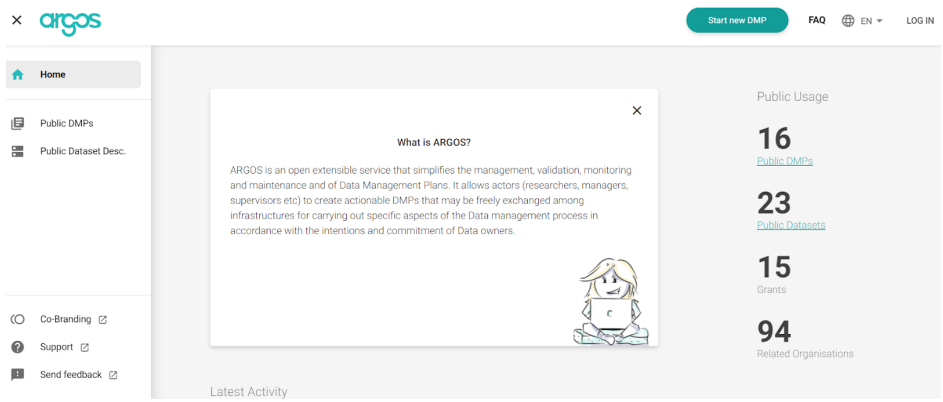

Figure 40: DMP section and Dataset section creation on the Argos tool

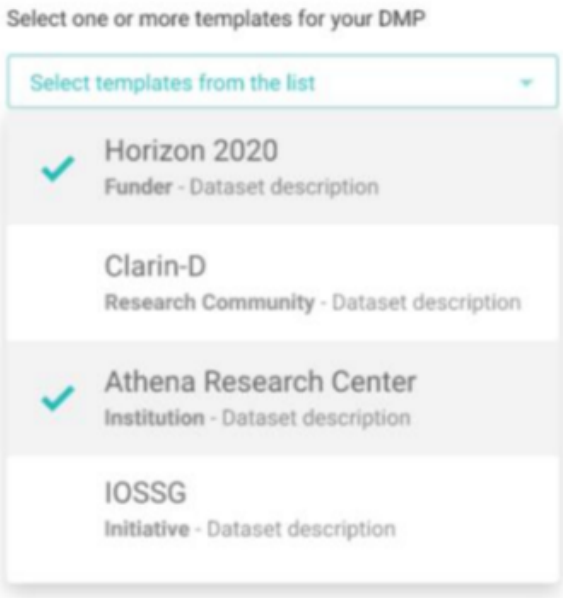

Figure 41: Using of the different templates for DMP creation

includes all-important descriptions related to the project when a section of the datasets provides information about data and their management [192, 195, 81, 196, 177].

In the same way, as DMP created on DMPOnline, DMP (maDMP) on Argos is created in private mode by default and can be opened when researchers need it and shared with colleagues for collaborative creation. Plans created on Argos are living, versioned documents, which can be updated at any time of the project. Moreover, the plans can be validated, for example, by institutional data steward for quality control. In the end, the plans can be published directly on Zenodo with DOI assignment and citation [192, 195, 81, 196, 177, 198, 197].

A DMP created in Argos is a rich document, discoverable through OpenAIRE, covering all RDM aspects of the project that may include one or more descriptions of the datasets [197, 198]. During the DMP/maDMP creation, the researcher can use more than one template, (Figure 41) easily choose OpenAIRE and EOSC resources, and share plans with colleagues. Argos allows sending a notification to repository managers, analyzing the usage statistics, defining indicators, exporting DMPs in JSON format [154, 208].

In administration mode (Figure 42), templates can be customized according to researchers' needs, additional functionality can be managed, APIs configured and RDA compliance controlled [195, 196].

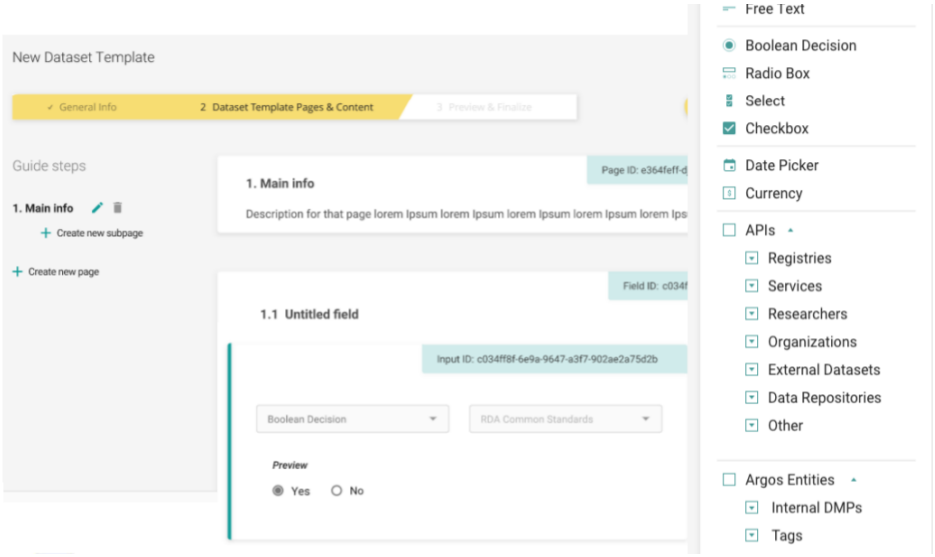

Figure 42: Customization of the templates in administration mode

Argos is being disseminated in different European countries through the network of national secretariats of OpenAIRE, through training sessions, and it is still contributing to its development by sending feedback based on users' experiences. The principal adopters of the Argos are Foundations for Science and Technology, both Spanish (FECYT) and Portuguese (FCT). The multilanguage interface allows having more users from over the world [208].

The long-term vision of Argo is based on the principles of maDMP and interoperability. In the near future, Argo's developers intend to implement features related to the execution of information in maDMP (semi-automated or automated realization), with the publication of data in the specified repository, with a certain license, for example. Validation of information in maDMP with FAIR, funder, and RDM requirements, as well as mechanisms for notifying authors or approving their plans, are also among the future directions [197].

With Argo's tool development, OpenAire highlighted the value of the maDMP, due to the possibility of exchanging information and the connection between different RDM services, which can not be executed in traditional DMP [197].

3.3.6 Other projects

During the establishment of the maDMP standard, the application profile was also tested in other tools. According to Miksa T. et al. [171], one of the applications was carried out at the Technical University of Vienna[101] during the course of Digital Preservation[102]. Series of the experiences showed that tools and services based on maDMP can help to select the more adequate repository, reuse existing information from other systems like ORCID, GitHub, and pre-fill DMP, identify the size, and the format of the data files, and, consequently, provide storage estimates. In other words, maDMP can diminish researchers' effort, automate the DMP creation process and improve the quality of the plan [171].

Another implementation of the maDMP occured in the context of the Austrian COMET K1 program and focused on the creation of the prototype of a maDMP on a content management system. As the result, the authors Bakos, A., Miksa, T., and Rauber, A. proved that a data management system could be extended with preservation functions where maDMP can help, for example, to attach the license to a dataset automatically. That, in turn, can help to reduce the researcher's workload

101 https://www.tuwien.ac.at
102 https://goo.gl/V6Zx3n

during RDM activities, making sure that their projects and data are in compliance with the funder policies [12].

According to Austin C., Kalonji F., Timms K., and Cuffe J. the maDMPs have a positive impact in a government context, helping to manage an extensive amount of records and datasets following the legal and policy requirements. They proposed an extension of the RDA maDMP and its logical model, which includes more details in legal and ethical issues, namely indigenous considerations, security, privacy, and intellectual property aspects, among others [216].

The F1000 Research platform[103] also implemented maDMP application profile with the aim of connecting and tracking information about researchers' projects for funders and institutions, improving interoperability between RDM systems and reducing effort [177].

The Norwegian Center for Research Data[104] shared experiences that by adopting maDMP on their side, they could ease the sharing of research data, providing recommendations for collecting, storing and preservation of the data, for example, suggesting repositories more adequate for research data by APIs from Re3data[105], using information indicated on maDMP.

The easyDMP tool[106] through maDMP implementation could improve the ability to reserve space on the storage services requested in the DMP.

Another proposal for improving the DMP creation process is related to the automation of feedback during the preparation of the grant applications. Based on experiences with the "Impacter" platform[107], Lefebvre A., Bakhtiari B., Spruit M. suggested "automated feedback technology using natural language processing techniques", which allows researchers to make their DMP more complete and improve the quality [154].

The list of the maADMP adapters[108] is constantly growing and the number of the RDM systems standardized with the maDMP standard is growing as well [175]. The RDA Hackathon on maDMPs[109] which included 89 participants from 21 countries showed that there are many projects seeking to contribute to the activities related to the maDMP. Different teams with different approaches and ideas were working on the achievement of the principal goal - taking more benefits from DMP while being divided into four topics: serialization, integration of DMP tools, funder template mapping, and further integration [34]. Creation of a new version of the DMP Common Standard Ontology [35], improvement of the code of the existing tool according to maDMP standard [181, 91, 251, 262, 149], mapping the maDMP application profile to funder templates [36, 219], and analysis of the existing institutional RDM workflow and systems for its improvement according to maDMP standard [286, 111, 269, 141] were the principal themes of the Hackathon.

Despite this, free text still cannot be avoided in a maDMP and a human-readable narrative is still needed [171]. MaDMPs make it easier to continuously and systematically monitor a DMP from the start to the end. The community of the maDMP adapters is growing, increasing the support for RDM systems based on maDMPs. MaDMPs are seen as means to exchange DMP information in a machine-actionable way, in multiple contexts, and for all stakeholders: researchers, funders, repository managers, research administrators, data librarians, etc., help to save time and reduce costs while providing more precise information [34].

103 https://f1000research.com/
104 https://www.nsd.no/en/
105 https://www.re3data.org/
106 https://easydmp.sigma2.no/
107 https://impacter.eu
108 https://activedmps.org/
109 https://www.rd-alliance.org/rda-hackathon-madmps

3.4 DOMAIN DATA PROTOCOLS

Due to the increase in the importance of the DMP creation, funders, research organizations, and universities are developing systems that will make this process easier. Moreover, they are developing systems that will allow accessing DMPs, understanding, comparing, and evaluating them. Science Europe recognizes the role of the Open Science policies, requirements, and practices. However, it confirms that there are many different templates for DMPs that vary significantly. In this context, they propose a framework for "disciplinary research data management protocols", called Domain Data Protocols [73].

The principal goal of the Domain Data Protocols is based on researchers' specific needs and allows them to have more adequate templates for their plans in any scientific domain. This also helps all RDM stakeholders to support RDM activities easier and to obtain more benefits from RDM in general [73].

More specifically, each scientific domain community can use the proposed Framework for the Domain Data Protocols' creation within their disciplinary needs. Complying with the Domain Data Protocols, researchers will have more adequate DMP for their domain. From the researchers' point of view, it will make the creation of a DMP easier and quicker, reducing the number of errors, Domain Data Protocols are "pre-filled DMPs"[110]. In turn, from the funders' point of view, it will simplify the DMP evaluation process by reducing the review time and different types of plans. In general, Domain Data Protocols can be seen as a possibility to improve the quality of the DMPs and help to standardize DMPs according to different scientific domains. Furthermore, this approach will simplify the DMP creation process and its evaluation, turning DMP from a "bureaucratic imposition" to a "win-win" document for all parties [73].

Domain Data Protocols will also meet RDM requirements, but with more attention to relevant standards and practices for RDM in specific domains. Furthermore, the creation of the Domain Data Protocols should be in compliance with FAIR principles, applicable standards, laws and regulations, have formal minimum conditions, support resources, templates and examples [73].

This approach is not mandatory yet but continues to be a recommendation [73] and can be compared with the existing Controlled Vocabularies approach for data and metadata from different scientific domains. This is because controlled vocabularies also have similar goals: improve, facilitate, and standardize the description of the data without increasing time and effort applied by researchers [133].

Some of the domains have already started to develop Domain Data Protocols, models, and templates for DMP creation in their domains, using controlled vocabularies [73].

Thus, for example, in the domain Humanities, Archeology during the project Parthenos/Ariadne, the creation of the template for a Humanities DMP has already started which is in line with the Domain Data Protocol [73, 101]. The Common Language Resources and Technology Infrastructure (CLARIN)[111] also demonstrated their willingness to co-operate with Science Europe in the development of Domain Data Protocols for its domain. CESSDA (Consortium of Social Science Data Archives)[112], ELIXIR (intergovernmental organization related to the Life Sciece domain)[113] evaluated Domain Data Protocols creation as very helpful in quality control [73, 29].

Controlled Vocabularies in the DMP can be related to the type of data, their size, formats, tools, methodologies, and software used during the project, among others. Thus, for example, in Environmental Data Management Plan for the Environmental Restoration Program created in 1996 was indicated different types of projects that

110 https://ddp-bildung.org/
111 https://www.clarin.eu/
112 https://www.cessda.eu/
113 https://www.elixir-europe.org

can exist for this domain: surveillance and maintenance, decontamination and decommissioning, remedial design/remedial action, and remedial investigation/feasibility studies. This list can be used in Domain Data Protocols as one of the blocks of the describing information about the project [119].

Authors Mischo W. H, Schlembach M. C, O'Donnell M. N. analyzed different DMP in their works to define controlled vocabulary terms that can be used in DMP. Most of the controlled vocabularies are related to the type of data, storage, location, licenses and access rules. For example, they have the following values for storage: computers, servers, hard drives, workstations, laboratory server, group computer, flash drives, etc.; for repositories, the following values were used: GenBank, arXiv, ICPSR, Dryad, Protein Data Bank, etc. Controlled vocabularies for funded status, for departments, grant numbers were also proposed to facilitate the creation of DMPs, making them more standardized [179].

Another example of the incorporation of the controlled vocabularies into DMPs was described by Stoddenn V., Ferrini V., Gabanyi M., Lehnert, K., Morton J., and Berman H. [250]. Using the ezDMP tool[114] and controlled vocabularies, researchers can create well-detailed DMP with minimal additional effort, resulting into documents with structured information allowing the machine-readability and exchanging. Controlled vocabularies were proposed and implemented in different fields of the ezDMP tool. For example, for choosing research products, five values were proposed: software, data products, curriculum, physical specimens and workflow information; for NSF submission target: Directorate for Biological Sciences, Directorate for Mathematical and Physical Sciences, etc. A list of the different research data repositories was also presented as a controlled vocabulary. According to the authors, the next generation of the DMP should include maDMP principles, controlled vocabularies, and semantic descriptions, which allow facilitating DMP creation, resulting in higher quality documents with structure information than can be easily reused in different RDM systems [250].

The Model Data Management Plan Standard with different terms, which in turn are parts of the controlled vocabularies, were created for Clinical Data Management Community. This model can be used for DMP creation for small and midsize pharmaceutical companies or large companies related to the clinical practices and data and includes different terms. For example, for a description of the staff related to the RDM, there are the following terms: data managers, database developers, programmers, project managers, quality assurance staff, quality control staff, staff involved in the creation of CDISC SDTM files (which is very specific for this domains), and sponsor/client [26]. Case report forms, electronic report forms, electronic data capture applications, testing the electronic data capture application, statistical analysis, clinical trial database, etc. are terms that also can be used as controlled vocabularies in DMP for this domain.

Based on the Domain Data Protocols approach proposed by Europe Science the RDA Discipline-Specific Guidance for DMPs Working Group was created with the aim to "propose discipline-specific adaptations for DMP templates for the disciplines behavioural, educational and social science, natural and engineering science as well as life science with the focus on medicine and biology". The work of this group is focused on differences of the certain disciplines' needs, related to the research data management issues. They also seek to develop discipline-specific recommendations and guidelines for DMPs, present them to the research community and implement them into different DMP tools [112].

Although their work is still under development, the RDA Discipline-Specific Guidance for DMPs Working Group has already created a survey "About Discipline-specific Guidance for DMPs"[115] to collect information related to the specific terms and differences that existed in scientific domains during the DMP creation and to create a catalog for different disciplines. They also analyzed the current state of

114 https://ezdmp.org/
115 https://www.soscisurvey.de/discipline-specific_guidance_dmp/?q=DMP

disciplinary-specific data management practices, knowledge or service gaps, and overlappings between disciplines [112]. The survey showed that there are many different terms that can be used for Domain Data Protocols creation. These terms can be seen as controlled vocabularies in DMPs. For example, they identified the following terms for the description of the type of data collected during the project: tabular data, textual data (such as interview transcripts or survey responses), references (i.e. bibliography), binary or numerical data, big data (heterological data types), audio recordings, visual recordings, audiovisual recordings, images, 3D models, algorithms, simulations, code/software, paper or digital documentation, physical samples, etc. Observations (human, instruments, or sensors), surveys, interviews, lab experiments, field experiments, crowdsourcing, web scraping, etc. were identified for the collection method[116].

Different workshops (e.g. eScience Days 2021 – Workshop "Discipline-specific guidance on data management plans") and dissemination sessions carried out by the RDA Discipline-Specific Guidance for DMPs Working Group besides the needs collection, help to demonstrate the importance and strengthen interest in Domain Data Protocols by other institutions [215].

Thus, based on Domain Data Protocols proposed by Europe Science and WG group work, the Domain Data Protocols for educational research in Germany, which includes twelve partnering institutions, are developing "standardized data protocols to ensure data quality and the re-use of research data"[117]. The Leibniz Institute for Astrophysics Potsdam[118], the German Institute for Adult Education, Leibniz Centre for Lifelong Learning[119], the Leibniz Institute for Research and Information in Education[120], the German Institute for Economic Research / German Socio-Economic Panel Study (SOEP)[121], the German Youth Institute[122], the German Centre for Higher Education Research and Science Studies[123], the Leibniz Institute for the Social Sciences[124], the Institute for Educational Quality Improvement[125], the Leibniz Institute for Educational Trajectories[126], the Qualiservice / University Bremen[127], the Center for Teacher Training and Education Research[128], the Leibniz Institute for Psychology[129] are the members of this group[130].

3.5 SUMMARY

To sum it up, the well-organized data during the project and its completion help to increase the value of the data, making the research project, researchers' effort, work and funds worthwhile. In this context, a DMP can help to structure and well-define the course of the project and their RDM activities. If it is kept up-to-date, it can be integrated with other systems and workflows (e.g., a data repository, grant, or project management system), and can lead the entire data strategy of a project. However, DMP creation, the same way as most RDM activities, is a time-consuming task that requires some effort, specific knowledge, some data publication experience, and the appropriate tools. This is why many institutions are looking for solutions to help researchers with creating DMPs and with RDM activities in general.

116 https://www.soscisurvey.de/discipline-specific_guidance_dmp/?q=DMP
117 https://ddp-bildung.org/about-the-project/
118 https://www.aip.de/en/
119 https://www.die-bonn.de/default.aspx?lang=en&
120 https://www.dipf.de/en?set_language=en
121 https://www.die-bonn.de/default.aspx?lang=en&
122 https://www.dji.de/en/about-us.html
123 https://www.dzhw.eu/en/index_html
124 https://www.gesis.org/en/home
125 https://www.iqb.hu-berlin.de/
126 https://www.lifbi.de/LIfBi-Home
127 https://www.qualiservice.org/en/
128 https://www.uni-potsdam.de/en/zelb/
129 https://www.leibniz-psychology.org/en/
130 https://ddp-bildung.org/about-the-project/

In this chapter, we describe in detail what a DMP is and what the content of a good DMP should have. We also explain the importance of a DMP, its benefits, and the DMP requirements that exist in the scientific community. We demonstrate various examples of DMPs created over the last decades, which despite their slightly different name, were as focused on improving data management during the project lifecycle as the current ones. In this chapter, we also show that DMP help in improving planning and organization issues in advance, avoiding undesirable unforeseen events and predicting possible problems at the beginning of a project with sufficient time to "fix" them.

Furthermore, this chapter covers machine-actionable DMPs, and the benefits of their development, such as the ability to make DMPs reusable by humans and machines or to improve the exchange and interconnection of RDM and DMP support systems. We look at various projects and tools to support researchers in creating DMPs and maDMPs, as well as various mechanisms for automating the DMP-related process. The huge efforts already made in this area are also highlighted by Data Domain Protocols aimed at standardizing and simplifying the DMP creation process, and by Controlled Vocabularies, which can also be useful in DMPs.

We show that there are many efforts in the scientific community to help researchers with RDM and DMP issues and that institutions understand and recognize the importance of a well-organized support system, developing various tools based on recommendations. In this context, the following chapter provides a systematic overview of the different ways of organization of institutional DMP support, the challenges that researchers face and the ways to overcome them.

4 INSTITUTIONAL SUPPORT FOR DMP

The importance of RDM, including the DMP's creation, is growing [145]. Some funding agencies require the inclusion of the DMPs in grant applications, while others require the plan within the first months of the project grants (e.g., projects under Horizon2020 [44], National Science Foundation [97])[1]. In order to comply with these requirements, institutions have been developing different ways to support researchers with creating DMPs and other RDM activities during the lifecycle of projects. They focus on both the engagement of researchers and the establishment of RDM infrastructures according to their needs. Thus, institutions can organize workshops and training sessions for researchers or develop and implement services and tools into their RDM workflow. The common goal is to support researchers in all RDM activities [75, 279, 39]. In this chapter, we present a systematic review related to the different ways DMP support exists in institutions, libraries, research centres, and other support departments and how they have been developed, implemented, and organized. This review helps us to analyze the strengths and weaknesses of various approaches and apply best practices in the DMP support system we are developing.

4.1 SAMPLING PROCEDURES

For the systematic review analysis, an approach based on the PRISMA 2020 statement[2] was adopted. This approach includes various activities such as the definition of the objectives and research questions for analysis, literature source, inclusion and exclusion criteria, analysis and discussion [194]. In short, the approach helps to collect studies related to the analyzed topic from different sources (e.g., Scopus database), apply inclusion and exclusion criteria, eliminate duplicated registers, and review more relevant and more important papers (Figure 43).

Therefore, in order to achieve our main objective in our systematic review, namely to identify global studies on DMP support in different institutions and understand how institutions provide, implement, and organize DMP support for researchers, we defined the following ***Research Questions for a Systematic Review (RQaSR):***

RQaSR1. What kind of DMP support do institutions provide for researchers?
RQaSR2. How is this support developed, implemented, and organized?

According to the PRISMA approach, first, we defined the Scopus Database[3] as a literature source for analysis to answer the RQaSRs.

Since DMP creation is a part of RDM activities, DMP support is a part of RDM in the institution. Therefore, our search is focused on the literature review base related to RDM in general. This includes the review of RDM services and RDM practices, which could be used in institutions to support researchers because literature describing the development of RDM support and services may include information about the development of DMP support and be a part of the development phase of RDM infrastructure in general. Moreover, we analyzed literature related to DMP support services, systems, and other types of support existing in the institution. Therefore, "research data management", "research data management practices", "research data

1 https://www.dcc.ac.uk/resources/data-management-plans/funders-requirements
2 http://www.prisma-statement.org/
3 https://www.scopus.com/

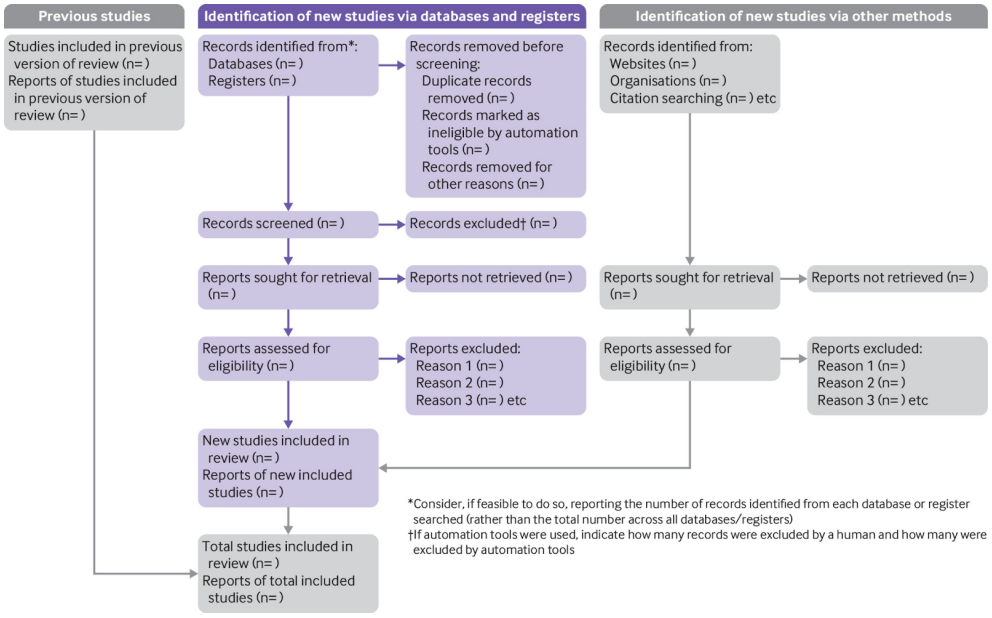

Figure 43: PRISMA 2020 flow diagram template for systematic reviews [194]

management services", "research data management support", "data management plan support", "data management plan support service", "data management plan system", "DMP support", "DMP support service", "DMP support system" were defined as key search terms and concepts. The search was conducted on 14 May 2022 in the Scopus Database, with the use of the following search strategy:

1. The first search query is a global search that included Title, Abstract, and Keywords:

TITLE-ABS-KEY (("research data management") OR ("research data management practices") OR ("research data management services") OR ("research data management support") OR ("data management plan support") OR ("data management plan support service") OR ("data management plan system") OR ("DMP support") OR ("DMP support service") OR ("DMP support system"))

As a result of this query, we have received **864** documents.

2. In order to limit our search, we decided to refine our query using only the field keyword focused on research data management issues:

(KEY ("research data management") OR KEY ("research data management practices") OR KEY ("research data management services") OR KEY ("research data management support") OR KEY ("data management plan support") OR KEY ("data management plan support service") OR KEY ("data management plan system") OR KEY ("DMP support") OR KEY ("DMP support service") OR KEY ("DMP support system"))

The result of this query was **542** documents.

3. To these query results, we applied the "Timeframe" filter, which included only the last five years, from 2018 to 2022. The reason for the range was to analyze the most recent and current articles and documents in order to make systematic review and literature reading more feasible for the context of the work being developed. In addition, as explained in Chapter 3, the subject related to the creation of DMPs is quite recent. Although other types of plans, helping in data management of projects, already existed in the past decade, they were not used for research projects in institutions. According to the literature, DMPs in the context of research data projects started to emerge in 2005 with requirements from funders.

(KEY ("research data management") OR KEY ("research data management practices") OR KEY ("research data management services") OR KEY ("research data management

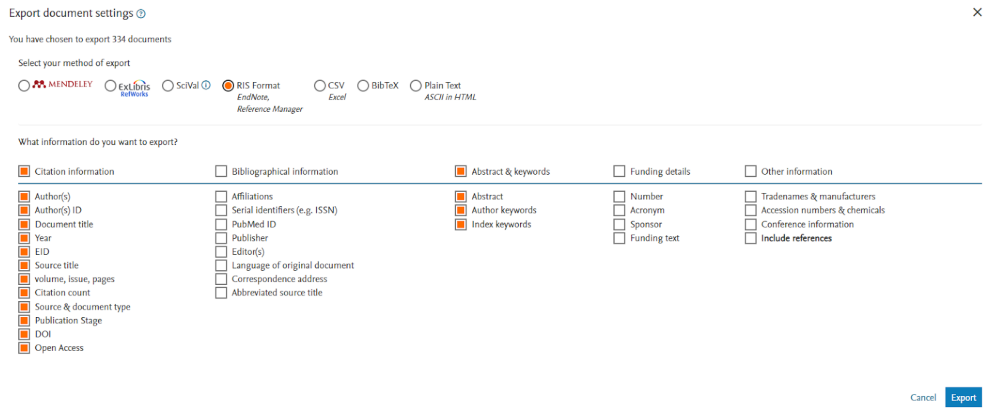

Figure 44: Records export to EndNote

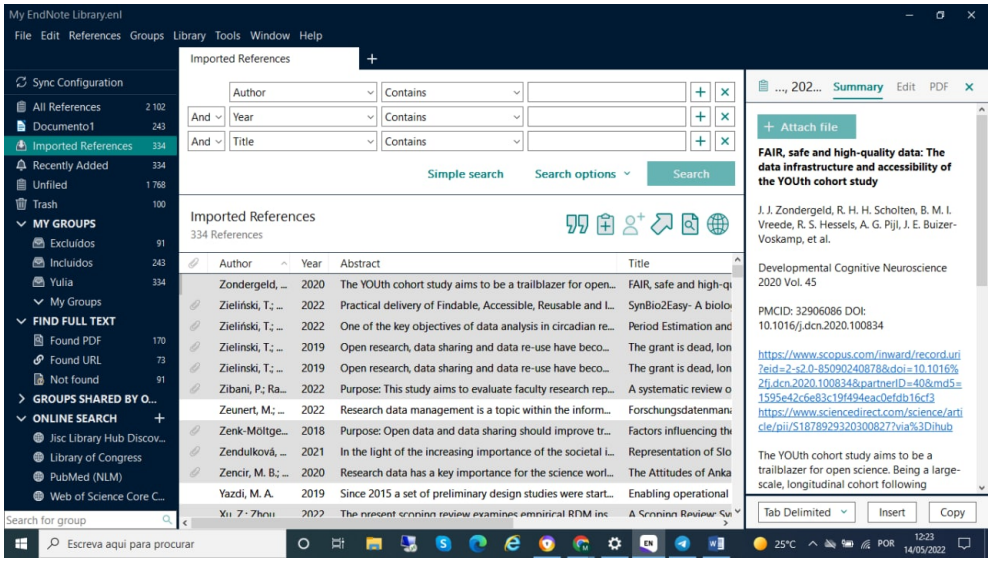

Figure 45: EndNote Interface with exported records

support") OR KEY ("data management plan support") OR KEY ("data management plan support service") OR KEY ("data management plan system") OR KEY ("DMP support") OR KEY ("DMP support service") OR KEY ("DMP support system")) AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018))

The "Timeframe" filter has reduced the number of documents to **334**, and the records have been exported to the EndNote reference management system, which has helped us organize, cite and format the references, continuing the review and analysis (Figure 44, 45).

4. Our document analysis was based on the abstract reading. However, as we started reading, we realized that it is not always possible to know if the article contains information about DMP support. This is because sometimes abstracts describe all types of support without separating DMP from other types of support. In case the analysis of the abstract was not conclusive, it was necessary to analyze the full text of the article. For this reason, with the EndNote option "FindFullText", we automatically excluded articles without access to the full text, thus reducing the result to **243** documents. At the same time, we checked for duplicates (Figure 46) and exported the result records to an Excel file, which was converted and deposited as a dataset "Bibliography and analysis on studies of institutional DMP support services" on the INESC TEC RDM repository [138].

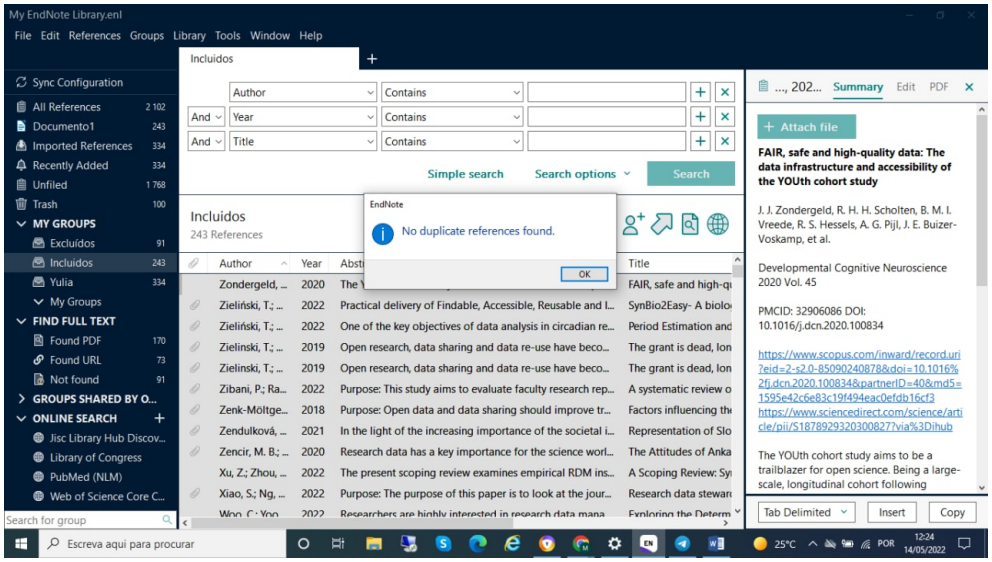

Figure 46: "FindFullText" application and verification of the duplicates

5. To proceed with the document analysis and answer our defined Research Questions for a Systematic Review, we defined exclusion and inclusion criteria.

Inclusion criteria:

– Studies that cover different ways of supporting researchers in DMP creation and monitoring processes.

– Studies that describe the development, configuration, and operation of DMP support systems and services in institutions, libraries, research centres, or IT departments.

– Studies that describe another type of support for researchers related to RDM issues, such as data deposit and description in the context of the DMP creation.

Exclusion criteria:

– Studies that do not describe DMP support issues.

In general, applying these criteria to the analysis, the final number of documents would include studies that describe different ways to support researchers in tasks related to the DMP creation and monitoring processes. These can be described as organized services, systems, and platforms developed and implemented, as well as other activities, such as workshops, webinars, and consulting by support departments. Documents describing support for researchers in other RDM tasks (e.g., data deposit, description) in the context of creating DMPs can also be included for consideration. Moreover, studies that describe the development of the support in other RDM tasks can also be included, if they show the institutions' efforts in meeting the RDM requirements regarding RDM services and support, including DMP requirements.

Therefore, the studies during the analysis of the application of the inclusion criteria were marked with "Yes", "No", or "Maybe", where "Yes" indicates studies that meet all three inclusion criteria, "No" indicates studies that do not meet all three inclusion criteria, and "Maybe" indicates studies that are questionable and require more detailed analysis. The column "Full Text Reading is required" with the options "Yes" or "No" shows whether the document is selected for further analysis of its full text.

This part of the analysis occurred in a double review process together with the research data librarian who supports researchers and students at the Nursing School

of Porto (ESEP)[4]. The different perspectives of the two researchers, given their professional backgrounds in RDM user support, in the dual review process, provide a more robust analysis.

After that, all selected documents were checked by exclusion criteria during the full-text analysis and marked as "Describes", "Not Describes" or "Maybe", where "Describes" are studies that contain a description of the DMP support, "Not Describes" are studies that have no description of the DMP support, the development of a support system, or services, and "Maybe" are studies that have a good detailed description of the RDM support services, their organization, development, and improvements that do not speak about the DMP support specifically but show the work for researchers support in general.

Therefore, after the inclusion criteria, the results of the studies were reduced to **56** and after the application of the exclusion criteria - to **14**. Author, Year, Title, Source, DOI, Keywords, Abstract, Institution, Inclusion criteria, FullTexReading, Exclusion criteria, Selected, and Comments are columns that were filled during the analysis on the Excel file that was deposited as a dataset of this systematic review [138].

6. In line with the PRISMA approach, we can add studies from other sources to the analysis. Therefore, for the broader coverage of the studies and in the context of our ongoing collaboration with various groups from the Research Data Alliance and our assumption that there is grey literature describing DMP support that has not reached pre-publication, we sent a request to the RDA. In the request, by email, we asked researchers, professors, data managers, data stewards, librarians, and other interested parties in RDM infrastructure and researcher support to indicate or send documents describing the development of support systems. Those could range from technical reports to books that, for whatever reason, have not been included in the Scopus database.

As a result of that collaboration, we received **12** papers, which we added to our Excel file for the application of the inclusion and exclusion criteria and **9** links for sites with different guides and practices used by RDA members:

1. Data management planning support in the University of Manchester Library Library[5]. Libraries can review DMPs and provide feedback within approximately 10 days if the researcher sends a request. Moreover, they provide different examples of the DMPs, guidance, and resources that can help researchers in DMP creation since a DMP is mandatory. For DMP creation, they use the DMPOnline tool;

2. FAIR Data Resources[6] - a collection of documents on Zotero, including documents for researchers, service providers, developers, and science managers who are interested in FAIR data management. Although FAIR principles are made as a part of the RDM and DMP, they do not describe the development, organization, or operation of the support for researchers;

3. FAIRsharing[7] - is an online resource that collects essential information to make data consistent with FAIR principles. This site presents standards, databases, policies and collections; however, it does not describe the development of researcher support as a system or service;

4. EOSC-Life[8] - European Open Science Cloud platform that provides resources to help researchers make their data FAIR and receive RDM support; however, it is still under development;

5. Digital Library Services at the University of Cape Town[9]. Data stewards at this University can help researchers create or review their DMPs, and provide templates, checklists, guides, and tutorials for DMP creation. Moreover, they provide other RDM support such as digital preservation, and digital curation, among others;

4 https://www.esenf.pt/pt/
5 https://tinyurl.com/kh6ynjxc
6 https://www.zotero.org/groups/2345721/fair_data_resources/
7 https://fairsharing.org/
8 https://www.eosc-life.eu/
9 http://www.digitalservices.lib.uct.ac.za/

Nº	Author	Year	Title	DOI	Keywords	Abstract	Exclusion Studies that not describe DMP support issues	Selected?	Comments
1	J. J.	#	FAIR, safe and high-quality data: The data	10.1016/	Cohor	The YOUth cohort study aims to be a trailblazer for open science.	Maybe	Y	FAIR, data management development:
2	S.	#	Research data stewardship at the University	10.1108/	Data	Purpose: The purpose of this paper is to look at the journey and	Describes	Y	DMP support, RDM support, DMP input
3	T.	#	Research data management and the	10.1835	Data	This case study critically examines ongoing developments in	Describes	Y	RDM and DMP support; contact by
4	V.	#	Establishing RDM services for small-scale	10.1835	Acade	During an international library conference in 2017 the authors had	Describes	Y	Support in DMP creation; support
5	M.	#	Who does what? - Research data	10.5334/	Data	We present the approach to Research Data Management (RDM)	Describes	Y	1 to 1 consulting, workshops, trainings,
6	J.	#	Expanding the research data management	10.5334/	Conti	Research Data Management at Bielefeld University is considered	Describes	Y	RDM support, DMP support
7	D.	#	Leading fair adoption across the institution: A	10.5334/	Drexel	Universities strive to foster knowledge sharing and greater	Maybe	Y	RDM support
8	H.	#	Research data management implementation	10.1162/	Data	Research Data Management (RDM) has become increasingly	Maybe	Y	RDM support description, RDM
9	A.	#	Exploring research data management	10.1515/	Data	Research data management planning (RDMP) is the process	Maybe	Y	RDM planning, survey analysis, quality
10	Y.	#	Institutional Support for Data Management	10.1007/	Data	Researchers are being prompted by funders and institutions to	Describes	Y	DMP support methodology
11	M.	#	The role of CRIS's in the research life cycle. A	10.1016/	Archiv	In 2015, Radboud University (Nijmegen, the Netherlands) started a	Describes	Y	DMP Workflow on system; DMP
12	R.	#	The embedded research librarian: A project	10.1835	Funde	This paper presents new services developed by the Lille University	Describes	Y	DMP support; RDM support
13	M.	#	Policy needs to go hand in hand with practice:	10.5334/	Code	In this paper, we explain our strategy for developing research data	Maybe	Y	RDM policy, research support very
14	Tim	#	Data Curation Initiative University of Miami	https://p		Data curation - the management of data throughout its research	Maybe	Y	RDM support description and
15	Ana	#	Forschungsdatenmanagement an der ETH	https://d	Resea	Not only the increasingly complex methods in research, but also	Describes	Y	RDM and DMP support, organization,
16	Lor	#	Keep Calm and Fill in Your DMP: Lessons	10.2218/		Aligning with other funders such as Horizon 2020, the Swiss	Describes	Y	RDM and DMP support, organization,
17	Uni		Strategic Plan for Research Data Services at	https://h		Based on evidence of need from in-person campus stakeholder	Describes	Y	RDM and DMP support, organization,
18	N.	#	Current Trends in Research Data	10.3103/		Abstract: This article presents an analysis of policies, guidelines	Describes	Y	RDM and DMP support, organization,

Figure 47: Total of the selected studies for systematic review

6. List of the Research Data Services provided by the University of Cape Town[10] with a full description of the services, software used and links for them;

7. Research Data Access and Preservation Association[11] with various presentations from summits organized over the years focusing on institutional repositories, data citation, almetrics, data infrastructure, linked data, metadata, data use and reuse;

8. Link to the "Adapting a Closed Class to Open: Creating a Customizable Class Template to Teach Research Data Management Skills to Researchers"[12] used to introduce basic RDM and DMP skills to researchers;

9. Science ouverte France site[13], which presents reviews of various institutions in France related to DMP services in French only.

Although some links are useful in DMP creation, the information contained in them does not describe the development and implementation of support services in institutions. They are more information resources, which can help the person in charge of the DMP support service to improve their skills and find useful resources. Thus, these resources are not included in the systematic review. However, **12** studies mentioned by RDA members were added to the analysis with inclusion and exclusion criteria. After applying the inclusion criteria, only **6** studies were left. Then, after applying the exclusion criteria, only **4** studies were added to the final revision, together with **14** studies selected from the Scopus database. The total of our studies is **18** (Figure 47).

A summary of the preparatory sample proceedings for the analysis is shown in Table 1 and Figure 48.

4.2 CORPUS DESCRIPTION

In order to answer our research questions and to achieve our objective, namely, to identify different ways of DMP support, their development, implementation, and operation, after selecting the studies, **18** studies were selected for detailed analysis (Table 2). Each study was labeled with an abbreviation for ease of use in subsequent analyses.

The selected studies describe the different ways of the DMP researchers' support. Some of them contain issues of RDM support in general; others contain issues specific to DMP. Some describe the development and operation of the RDM infrastructure and organizational and technical aspects. Others concentrate on stakeholders and roles allocated in institutions to support researchers.

10 http://www.digitalservices.lib.uct.ac.za/dls/about/rdsfullservices
11 https://rdapassociation.org/past-summits
12 https://osf.io/TZWRN/
13 https://scienceouverte.couperin.org/retours-dexperience/

Table 1: Research design of the sample proceedings

Objectives	Identify global studies on DMP support in different institutions
Research Questions	1) What kind of DMP support do institutions provide for researchers? 2) How is this support developed, implemented and organized?
Search Date	14/05/2022
Equation search strategy (e.g., all fields)	TITLE-ABS-KEY (("research data management") OR ("research data management practices") OR ("research data management services") OR ("research data management support") OR ("data management plan support") OR ("data management plan support service") OR ("data management plan system") OR ("DMP support") OR ("DMP support service") OR (" DMP support system")) - n=864
Equation search strategy (e.g., key)	*(KEY ("research data management") OR KEY ("research data management practices") OR KEY ("research data management services") OR KEY ("research data management support") OR KEY ("data management plan support") OR KEY ("data management plan support service") OR KEY ("data management plan system") OR KEY ("DMP support") OR KEY ("DMP support service") OR KEY (" DMP support system"))* - *n=542*
Timeframe	5 last years (KEY ("research data management") OR KEY ("research data management practices") OR KEY ("research data management services") OR KEY ("research data management support") OR KEY ("data management plan support") OR KEY ("data management plan support service") OR KEY ("data management plan system") OR KEY ("DMP support") OR KEY ("DMP support service") OR KEY (" DMP support system")) AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018)) - n=334
FindFullText	n= 334 included n=243 excluded n=91
Databases	Scopus
Other sources	email for RDA group members related to the RDM and DMP - n=12
Software for analysis	EndNote for reference management Excel
Inclusion Criteria	- Studies that cover different ways of supporting researchers in DMP creation and monitoring processes. - Studies that describe the development, configuration, and operation of DMP support systems and services in institutions, libraries, research centres, or IT departments. - Studies that describe another type of support for researchers related to RDM issues, such as a data deposit and description in the context of the DMP creation.
Exclusion Criteria	- Studies that do not describe DMP support issues.
Methodological approach	PRISMA Statement (https://www.bmj.com/content/372/bmj.n71)
Data analysis	Abstract reading and Full text reading Double review process
Bibliography Software	EndNOTE
Results	Based on the readings and after applying exclusion and inclusion criteria RAW Data Scopus - 864 With limits defined Scopus - 542 With limits to a timeframe- 334 After text finder full text - 243 After inclusion criteria - 5616 After exclusion criteria - 14 RAW Data RDA - 12 After inclusion criteria - 6 After exclusion criteria - 4 Total studies - 14+4= 18

1. YOUth [287]: The paper was published in 2020. It describes data management infrastructure and the researchers' support for large-scale study YOUth, where data are collected by the Utrecht University[14] and the University Medical Center Utrecht[15]. They have full-time data managers who help researchers and the University to comply with RDM and open science requirements, FAIR principles, and improve infrastructure according to the needs. Moreover, data managers collaborate with the IT departments of the University and with the Utrecht University Library. During the YOUth project different types of data (e.g. personal, sensitive and private) are collected. That is why, a lot of attention is given to data security. Although the paper does not describe DMP support, we have highlighted the following impor-

14 https://www.uu.nl/en
15 https://www.umcutrecht.nl/en/

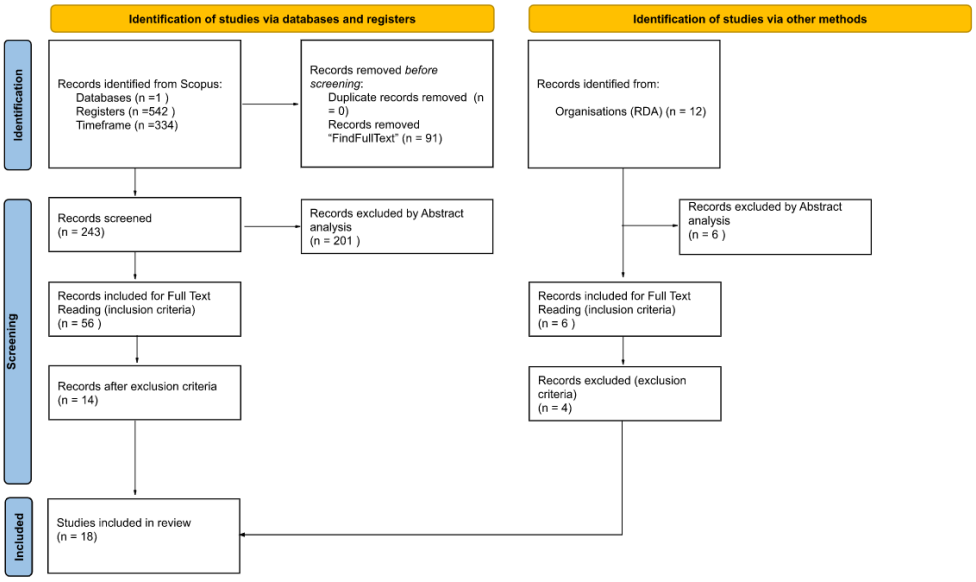

Figure 48: Flow diagram of the systematic review according to PRISMA approach

Table 2: Corpus of studies for analysis

Label	Title
YOUth	FAIR, safe and high-quality data: The data infrastructure and accessibility of the YOUth cohort study
HongKong	Research data stewardship at the University of Hong Kong
KU Leuven	Research data management and the evolutions of scholarship: Policy, infrastructure and data literacy at KU Leuven
Small-scale	Establishing RDM services for small-scale data producers at big universities
ETH Zurich	Who does what? - Research data management at ETH Zurich
ETH Zurich2	Forschungsdatenmanagement an der ETH Zürich: Ansätze und Wirkung
Three-pillar	Expanding the research data management service portfolio at Bielefeld University according to the three-pillar principle towards data FAIRness
Collaboration	Leading FAIR adoption across the institution: A collaboration between an academic library and a technology provider
Peking	Research data management implementation at Peking university library: Foster and promote open science and open data
RDM planning challenges	Exploring research data management planning challenges in practice
Five Case Studies	Institutional Support for Data Management Plans: Five Case Studies
Radboud	The role of CRIS's in the research lifecycle. A case study on implementing a FAIR RDM policy at Radboud University, the Netherlands
Partner	The embedded research librarian: A project partner
Policy needs	Policy needs to go hand in hand with practice: The learning and listening approach to data management
Miami	Data Curation Initiative University of Miami
Learn from a Swiss	Keep Calm and Fill in Your DMP: Lessons Learnt from a Swiss DMP-Template Initiative
Minnesota	Strategic Plan for Research Data Services at the University of Minnesota Libraries
Current Trends	Current Trends in Research Data Management

tant and valuable points related to the RDM issues and support in general, which show the motivation and interest in the development and improvement of RDM infrastructure and workflows:

1. Data managers interact with the researchers, data, and IT division daily, proactively working to improve all aspects of the data management and institutional support;

2. They focus on the collaboration with the researchers and with the partners' organizations, which also realize the importance of open science issues and efforts to improve RDM services;

3. They configured different types of storage: for biological samples - the Utrecht Biobank[16]; for tabular data - the UMCU Research Data platform[17]; for other types of data, including imaging data - the Yoda system[18];

4. Data managers also support the data transfer process, and when anomalies occur, they correct them to guarantee high-quality data. Moreover, to guarantee interoperability, data can be converted to a more appropriate format;

5. Description of the data is also supported by specialists of the University Library[19], using a script and different controlled vocabularies;

6. Regarding data access, they created a protocol with requests for internal or external researchers to guarantee data protection;

7. Finally, all RDM tasks are according to the series of the Standard Operating Procedures that were created and approved by the institution[20]; and all RDM staff are constantly organizing different events, training sessions, monitoring the quality of the data, and registration of the aspects that can be improved.

2. HongKong [285]: This paper, published in 2021, describes the development of the RDM infrastructure at the University of Hong Kong[21]. This work includes many activities focused on the Open Science policies and RDM requirements and answers the researchers' needs. The development of the RDM infrastructure was based on the collaboration of the Office of Knowledge Exchange[22], a division responsible for faculty members, of the University Libraries and, subsequently, of the Research Service Office, the Graduate School, individual faculties and of the division of Technology Services Support within the University Libraries created at the University[23]. The development of the RDM infrastructure resulted in the creation of the Institutional Repository HKU Scholars Hub[24], the Institutional Data Repository DataHub[25] and in the implementation of the RDM and DMP support services. Moreover, different guidance, web resources, institutional policies, and training programs were created at the University. Their RDM approach was based on the analysis of universities in the UK and Australia, with a principal focus on Open Science policies, RDM and funder requirements, and knowledge development and data literacy for staff. RDM infrastructure development started with institutional policy formation, which was indicated as an essential and fundamental component. Then the University implemented different support services, promoting an open science culture and engagement of the researchers at the same time:

1. For institutional policy development, a special group led by the University Research Committee was formed. It included representatives from different faculties, the Librarian, the Director of the Information Technology Services and the University Archivist for the development of a formal data policy for HKU. They analyzed the researchers' and the University's needs in the context of the Open Access and RDM requirements, including all aspects such as DMP, and data collection storage, preservation, sharing, and internal operational procedures (e.g., infrastructure, software, training sessions and support);

2. The development of the DMP support service was one of the principal points in the establishment of the institutional RDM policy. Different workflows for DMP creation and submission for researchers and postgraduate students and for funded projects were created. The DMP submission was mandatory, so all researchers with approved grants needed to create their plans. Moreover, data management should have been according to the information indicated on DMP. Thus, DMP should have been approved by the supervisor in case of the research postgraduate students or by

16 https://www.umcutrecht.nl/nl/centrale-biobank
17 https://dataverse.nl/dataverse/UMCU
18 https://www.uu.nl/en/research/yoda
19 https://www.uu.nl/en/university-library
20 https://www.uu.nl/en/research/youth-cohort-study/youth-standard-operating-procedures
21 https://www.hku.hk/
22 https://www.ke.hku.hk/about-ke/ke-office
23 https://lib.hku.hk/
24 https://hub.hku.hk/
25 https://libguides.lib.hku.hk/researchdata/datahub

the Head of Department/ Faculty Dean in case of the external grants. In this context, the University adopted DMPTool and created a webpage with an automated system for online DMP submission[26] and guidance for DMP creation[27]. After the DMP creation, it should have been approved by the Department Head after sending him an email notification automatically upon the submission on the "DMP Input Form". The plan could be approved or rejected, with automatic notification to the author. Currently, the HKU librarian and data managers provide different types of consultations related to the RDM aspects, such as individual, group training sessions, and consultations by email or telephone. One of the future work related to DMP support is the automation of the DMP creation processes, for example, costs issues;

3. Series of RDM training sessions, workshops, online resources, materials, videos, and advisory services helping researchers with RDM issues are also proposed by the HKU, for example, on the HKU RDS Website[28]. The Data and Scholarly Communication Librarian team supports researchers in all RDM aspects by e-mail and telephone line, individually by making appointments;

4. The HKU also pays a lot of attention to improving the knowledge and skills of the staff. They do not only recruit talents from external sectors but also provide different training sessions for current personnel. Workshops, courses, training activities and seminars are organized constantly. For example, the "Management of Research Data and Records" seminar is mandatory for newly enrolled or promoted academic and research staff.

3. KU Leuven [275]: The paper published in 2019 describes the effort that KU Leuven libraries[29], together with learning centres across Research Foundation Flanders[30] and the Research Coordination Office[31], had made for the development and implementation of the RDM policy, infrastructure, and support services at the institution. The development of RDM was included in the libraries' strategic plans and inspired by the Delft University of Technology[32] and Ghent University[33].

For researchers' support in the KU Leuven Library, a central RDM support desk was created. It provided first-line support and included library staff experts, information specialists and data stewards. They collaborate with the university's IT services, the Legal Department, KU Leuven Research and Development, the Research Coordination Office, and the library systems developer LIBIS[34]. The RDM support services were based on the RISE framework[35] which covers important RDM topics, such as RDM policy, strategy, advisory services, training sessions, DMP, risk assessment, preservation, and access issues, publication.

Recognizing the establishment of the institutional RDM policy as a principal task, they implemented it in 2014 with a principal focus on data publication issues. Then, constantly working to improve and make their RDM infrastructure according to the RDM requirements, a DMP user-oriented support service was proposed. Currently, the data steward is involved in planning the research project. They clarify all researchers' issues and analyze difficulties that researchers faced. In some cases, data stewards are cited as co-authors. In addition, support can be provided by email or ticketing system that was implemented in 2019 to attend to researchers' needs related to all aspects of the RDM, such as description and deposit.

Digital literacy is also highlighted in their work. It is important to develop training materials and workshops and improve the knowledge of both researchers and staff. They follow the "train the trainer" philosophy and prepare information spe-

26 https://dmp.lib.hku.hk/
27 https://hub.hku.hk/researchdata/staff10.htm
28 https://hub.hku.hk/researchdata/
29 https://www.arts.kuleuven.be/mosa/images/ku-leuven-library.jpg/image_view_fullscreen
30 https://www.fwo.be/en/
31 https://www.kuleuven.be/english/research/support
32 https://www.tudelft.nl/
33 https://www.ugent.be/en
34 https://bib.kuleuven.be/english/libis
35 https://www.digitalresearchservices.ed.ac.uk/resources/rise-framework

cialists and librarians for RDM support. RDM training was also implemented as a part of the Doctoral Schools of the Biomedical Sciences and Science and Technology groups and at the Faculty of the Arts. An undergraduate course on data and information literacy is also planned at the University. Moreover, RDM staff contribute to external activities, for example, in different RDM workshops or presentations at conferences, such as Open Belgium[36].

When introducing RDM support, they also pay attention to scalability, as there is no one-size-fits-all service. Therefore, the needs of researchers are analyzed all the time, gaps and weaknesses in the infrastructure are identified, and improvements are proposed.

4. Small-scale [268]: This paper, published in 2018, describes the RDM infrastructure of the two Universities: KTH Royal Institute of Technology Stockholm[37] and Westfälische Wilhelms-Universität Münster WWU[38]. Looking ahead, we can say that in both universities, RDM support services are under development. However, they provide a series of tools and guidance that help researchers in RDM activities. Data from KTH projects, for example, can be stored at the Swedish National Infrastructure for Computing[39], they have up to 200TB for 4 to 5 years of data preservation. Moreover, KTH created KTH Community on Zenodo[40] where researchers can also deposit their data without any limit. For DMP creation they adopted the DMPonline tool created by DCC and provided useful links for resources with guides[41].

Although they do not have well-organized DMP support, we can identify some positive points:

1. At KTH, there are staff from different scientific domains, so RDM support can be for multiple domain projects;

2. KTH Research Office[42] helps researchers to create a DMP during grant applications by an e-mail request, chat on-site, contact form, or phone[43];

3. They are motivated to develop and implement more specific services; however, awaiting for institutional formal mandate approved by the President of KTH. For now, informal support can be received from KTH libraries, archive staff, IT staff, Research Office, and the Sweden National Infrastructure Computing PDCentre[44].

The WWU, in turn, have already implemented the RDM Policy[45] and to support researchers they have a collaboration between the University libraries, IT services, and administration, calling "IKM" ("Information, Kommunikation und Medien" = information, communication, media). They focus on the development of the RDM infrastructure, training sessions, materials for support, e-learning, and storage. Moreover, at the WWU University, there is ULB - central literature and information-supply institution - a coordinator who helps in collaboration between researchers and support staff, and eScience Center[46] that provides support services and resources. However, as we mentioned above, DMP support services are being developed.

At the moment, there are three principal activities related to RDM at the University: 1) management of the research information system (CRIS), 2) management of the personal data, and 3) IT services related to technical issues. Libraries also help researchers with RDM tasks, such as data deposit, description, training sessions, workflows, and promotion of RDM literacy.

36 https://2022.openbelgium.be/
37 https://www.kth.se/en
38 https://www.uni-muenster.de/de/
39 https://www.snic.se/
40 https://zenodo.org/communities/kth?page=1&size=20
41 https://intra.kth.se/en/forskning/overgripande-stod/datahanteringsplan-1.814220
42 https://www.kth.se/en/forskning/research-office/research-office-1.76819
43 https://tinyurl.com/mvcpx58s
44 https://www.snic.se/
45 https://www.uni-muenster.de/Forschungsdaten/information/richtlinien/
46 https://www.uni-muenster.de/EScience/

Development of the research data repository, DMP tools and DMP support is on the list for future developments and will be based on the analysis of the researchers' needs.

5. ETH Zurich [260] and **6. ETH Zurich 2** [202]: The approach to RDM support at the ETH Zurich University[47] is described in their papers published in 2019 and 2020. The Human Resources department, the ETH Library[48], the ETH Zurich IT Services and the Research Data Management and Digital Curation Group at ETH Library play a central role in the development of the RDM infrastructure and focus on the DMP, active RDM and data publication and preservation issues.

Librarians support researchers from the beginning of the project, starting with advice for DMP creation, and helping with all RDM aspects. Collaboration with the Scientific IT Services also allows them to clarify researchers' issues related to the data repository, storage and preservation issues[49]. In this context, the information management system openBIS[50] was developed in 2007 to improve the RDM infrastructure of the University and help in several RDM activities, such as inventory management, management of physical samples, FAIRness data, annotation data, data analysis, and data audit, among others.

Regarding the DMP support, along with a DMP template[51], DMP checklist[52] providing for researchers, a series of training sessions and workshops that are carried out twice a year, to improve RDM knowledge and engage PhD students, researchers, and post-docs, the RDM summer school is organized. Librarians offer a "Book a Librarian service" to support researchers in all aspects related to the RDM, including DMP creation. The request can be made by email or personally. Moreover, the University staff also collaborate with the different departments, groups, and other institutions for tracking the know-how in RDM services and best practices, to apply them.

For data sharing, publication, and preservation activities, the RDM staff of the University encourage researchers to document their data from the beginning, clarify their questions, train them and provide different courses and guidance, for example, throughout the information website[53]. Moreover, they provide ETH Research Data Hub[54] for annotating, storing and managing all experimental and computational data at any stage and the Research Collection[55] for paper, thesis and other works publication.

For other activities related to the RDM, such as data visualization and data analysis, they offer services that can help to create and run data analysis workflows, analyze data, and code[56], among others[57], [58].

Although some RDM services and tools have been developed and implemented, the RDM infrastructure still needs refining. Institutional policies are still under development, with a focus on aligning the University with the requirements of RDM and funders, as well as with FAIR principles. The authors of the paper highlighted that the challenges of RDM cannot be solved by institutional planning and organization alone; they require national efforts, collaboration with other higher education institutions in Switzerland, and knowledge exchange on new solutions and best practices.

47 https://ethz.ch/en.html
48 https://library.ethz.ch/en/
49 www.ethz.ch/researchdata
50 https://openbis.ch/
51 https://sis.id.ethz.ch/services/rdm/SNSF-DMP-openBIS-template.pdf
52 https://www.dlcm.ch/download_file/force/66/275
53 https://sis.id.ethz.ch/services/rdm/
54 https://tinyurl.com/4584ap5s
55 https://www.research-collection.ethz.ch/
56 https://sis.id.ethz.ch/services/scientificcomputing/
57 https://sis.id.ethz.ch/services/scientificvisualization/
58 https://openbis.ch/

7. Three-pillar [230]: This paper, published in 2019, describes the three principles of RDM at Bielefeld University[59].One of them is policy development; the second one is the implementation of technical infrastructure, and the third is the establishment of support structures. The development of the RDM infrastructure started as a part of the Bielefeld Data Informium project[60] in 2010. Currently, it being expanded with the creation of the Competence Research Data Centre. In addition, collaboration with the library and computer center has helped the University to become more compliant with RDM and funder requirements, and implement RDM policy[61], provide central support for RDM, create a network with experts and collaborate with the external community and other institutions. In this context, the library has played a central role in the creation of various support services from the very beginning.

According to the authors, *"the University of Bielefeld is one of the first universities in Germany to adopt principles and guidelines on the handling of research data"*. They provide technical, consulting, and support services, both individual and group, seminars, workshops, training sessions and useful guidance that help researchers in RDM tasks[62]. RDM support includes consulting related to the DMP funder recommendations, policies, and legal aspects and can be provided at the early stage to avoid unforeseen situations in the research projects. Moreover, during their work, librarians contributed to the development of the Data Service Center for Business and Organizational Data, which in turn, gave the opportunity of the data deposit for universities researchers with backup's management and data visualization tools. Participation in projects, such as "From Heterogeneities to Inequalities", "Practices of comparing. Ordering and changing the world" and "Information Infrastructure", also shows the motivation of the university staff to follow the innovations related to the RDM.

At the moment, the following tools were implemented at the University: "DMP Tool"[63] based on the RDMO concept[64], cloud storage service "Sciebo" that allows storage, sharing and synchronisation of university-related data[65], GitLab for versioned and controlled storage of files and software[66] and the institutional repository of Bielefeld University "PUB – Publications at Bielefeld University"[67].

Improvement of staff knowledge is also important at the University. Different seminars and teaching courses are constantly organized in the University and outside of it.

Recognizing the importance of the RDM, the University Library, the computing centre "BITS", and the Bielefeld Center for Data Science[68] are developing the competence centre for research data focusing on the extension of the provided RDM and support services, improvement of the RDM infrastructure, and collaboration with national and international institutions.

8. Collaboration [188]: The establishment of the RDM infrastructure at the Drexel University[69] is described in the paper "Leading FAIR Adoption Across the Institution: A Collaboration Between an Academic Library and a Technology Provider" published in 2021. Its development started in 2015. However, without establishing a governance policy, the University librarians find the external collaboration to create RDM tools and support. In collaboration with the Compliance Officer, and the General Counsel Office, which included Chief Security, Privacy Officers, the Office of Research, IT staff, and leaders from the Office of Research, they organized a

59 https://www.uni-bielefeld.de/(en)/
60 http://129.70.43.116/de/informium
61 https://www.uni-bielefeld.de/ub/digital/forschungsdaten/policy/
62 https://www.uni-bielefeld.de/ub/digital/forschungsdaten/
63 https://rdmo.uni-bielefeld.de/
64 https://rdmorganiser.github.io/
65 https://sciebo.de/
66 https://gitlab.ub.uni-bielefeld.de/users/sign_in
67 https://pub.uni-bielefeld.de/
68 https://www.uni-bielefeld.de/einrichtungen/bicdas/
69 https://drexel.edu/

Data Stewardship Forum which provided materials, workshops, policies, information related to the RDM and support services. Since 2018, the University along with the Library has increased the effort regarding the RDM infrastructure and FAIR principles, hiring data stewards and experts in RDM.

They improved the Libraries management system according to the RDM and funders requirements[70]. Different consultation, training sessions, courses and guidelines, for example, for DMP creation[71], or DMP review are some of the support that libraries staff offer by email or in person.

Collaboration with the Ex Libris resulted in a research services platform to support RDM "Alma library management system"[72] that provides series of the technical and in-depth support. This collaboration was seen as a mutual interest and a benefit for both sides. Libraries receive the tools and providers familiar with the researchers' needs.

9. Peking [187]: The paper, published in 2020, describes a series of the activities that Peking University [73] carried out to develop its RDM infrastructure. Starting in 2010 the University implemented the Peking University Institutional Repository[74], Open Journals, and the Open Research Data Repository. With the introduction of RDM services in 2015, Peking University Library[75] has influenced the national open science policy by organizing various RDM-related events, workshops, seminars, and training sessions addressed to researchers, students, and staff. Moreover, the library staff is always looking for national and international cooperation opportunities to improve knowledge, keep abreast of know-how in the field, and continuously improve the RDM infrastructure.

One partner was the Social Science Research Institute, which played an important role in organizing workshops and training sessions. They also developed a data repository, which was managed by the University Management Research Data Centre, and collaborated with the Office of Scientific Research and the Social Science Research Office.

Another collaboration with the Bioinformatics Centre of Peking University[76] led to the connection of the bioinformatics database with the university data, a collaboration with the Beijing Information Resource Management Centre in linking the Beijing State Data Resource System and the National Information Centre in collecting valuable data across China.

Library and university staff provide RDM and DMP support via chat, email and phone. They also focus on analyzing the researchers' needs, conducting surveys to improve existing services. Furthermore, staff work on the establishment of the collaboration with the different research university units, administrative and other units on campus to invest and improve RDM support, creating an RDM framework.

Although work on the establishment and implementation of RDM services continues[77], the institutional RDM policy still needs improvements and is formally disseminated.

10. RDM planning challenges [154]: This paper, published in 2020, does not describe the DMP support at the institution. However, it describes many important issues, which should be addressed when developing and organizing DMP support, which has been identified thanks to the case study carried out in the Netherlands institutions.

One of the important aspects related to the quality of the DMP is its completeness, openness, and actualization that require support from the experts in DMP creation with more quality. In this context, it can be seen as advice for the organization of

70 https://www.library.drexel.edu/services/research-data-management-support/
71 https://www.youtube.com/watch?v=32ymiVoU5B8
72 https://www.exlibrisgroup.com/products/alma-library-services-platform/
73 https://english.pku.edu.cn/
74 https://ir.pku.edu.cn/
75 https://www.lib.pku.edu.cn/portal/en
76 https://www.cbi.pku.edu.cn/
77 https://www.lib.pku.edu.cn/portal/en/fw/sjfw/keyanshuju

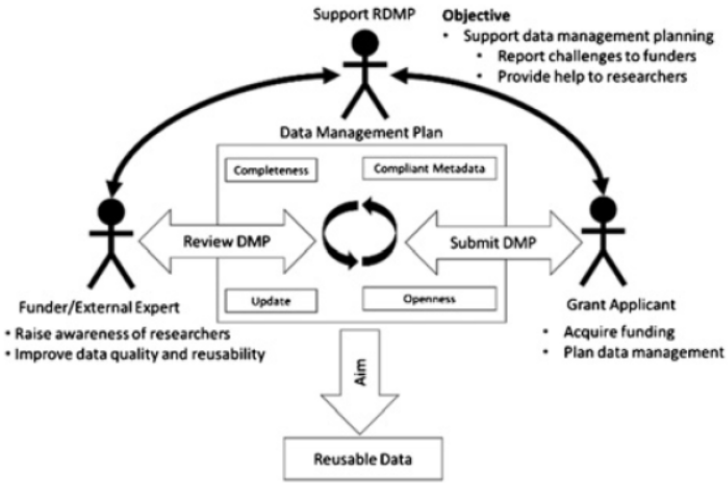

Figure 49: RDMP schema proposed by authors [154]

DMP support, based on the existence of experts to assist the researcher in answering any questions that may arise.

Another important aspect is related to feedback that can be provided by funders, and external and internal experts. The authors have proposed an RDMP scheme (Figure 49) in which DMP review, DMP support, and DMP submission are defined as the principal processes in RDM. Moreover, the author has confirmed that researchers have different types of difficulties in the DMP creation process, which require consulting support.

The third important point is the automation of the processes. They propose to integrate the automated feedback on the DMP creation process, for example, comments related to the data format during the creation of the DMP; "If your data do not have a common form, please provide a relevant convert tool".

To sum up, the authors have highlighted that RDM planning requires more attention, development of the quality criteria, and implementation of the best practices that could simplify the DMP creation, improving plans quality.

11. Five Case Studies [146]: This paper is our work. It was published in 2021. It describes a collaborative approach for supporting researchers during DMP creation. The approach was tested in RDM workflow at INESC TEC [78] with different scientific projects. It includes several activities, where a data steward assists researchers in all RDM aspects during the DMP creation.

The data steward started the collaboration by interviewing researchers to understand the context of the project. Then analyzes what type of data will be collected during the project, and, if will be required, additional documents such as Data Protection Impact Assessment. Moreover, information related to the data description, deposit, preservation, and organization were also identified, resulting in the first DMP draft created by the data steward. After that, researchers need to validate and improve the document if it is necessary. Finally, the DMP is published, for example, on Zenodo and a monitoring session is scheduled. In addition, the data steward supports researchers in data deposit, data description, and preservation providing access and assistance with different tools, such as the INESC TEC research data repository[79], the Dendro platform[80] and the LabTablet[81]. In some cases, collaboration with the DPO and IT staff is required.

Although this approach is the beginning of establishing DMP support, it shows that it is very useful for both researchers and the institution in general. Researchers

78 https://www.inesctec.pt/
79 https://rdm.inesctec.pt/
80 https://github.com/feup-infolab/dendro
81 https://github.com/feup-infolab/labtablet

receive a well-detailed published plan for their project, reducing risks and problems that may arise during the course of the project, as well as receiving expert support to monitor the plan and changes in the project, and training sessions on the DMP creation. In turn, the institute gains more compliant projects with RDM and funder requirements, better-organized data during and after project completion, and well-established RDM and DMP support, thereby raising its level among institutions. Moreover, this approach helps to identify lacks in RDM infrastructure, analyze difficulties that researchers face, suggest improvements, and implement different tools and establish collaboration both inside and outside the INESC TEC with other institutions. In more detail, this approach is described in the next Chapter 5.

12. Radboud [123]: In 2019, the authors Jettena, M. Simons E., and Rijnders, J. published a paper describing the implementation of the DMP module in the Radboud University Current Research Information Systems[82]. This module helps researchers create DMP, linking datasets of the project with the Dutch national DANS repository[83], the Radboud repository[84] and the Research Information Services to provide feedback for DMP. The support service desk structure plays an important role at the University, assisting researchers in all RDM aspects according to the institutional RDM policy during the data lifecycle of the project, such as data description, deposit, or storage.

To sum up, the Research Information System helps researchers to manage their projects, describe their datasets, deposit data on the DANS, link them with the publications, create DMP linked to the datasets and the project, collaborate with other researchers, and obtain feedback from support service.

The module DMP includes a series of questions that researchers have to fill in and it is based on the DMP Online tool[85]. It provides different templates from funders and guidance. In this case, researchers create a plan without any support, but once the plan is created, an email is automatically sent to the CRIS manager to verify the information and register the DMP. Then the researcher can request feedback by clicking on the appropriate button, automatically sending an email to the support team, receiving comments and suggestions on how to improve the document.

The authors of this paper show the importance of implementing the DMP module not only to help researchers create a plan and link data to the project and repository but also to demonstrate the university's compliance efforts to RDM and funder requirements by offering a wide range of services and tools for every RDM task in the data and project lifecycle. Moreover, the authors emphasize that the implementation of the DMP module contributes to FAIR compliance from the very beginning of the project, involving researchers in RDM activities and helping them to organize the data, making the project more efficient.

13. Partner [93]: The article "Embedded Research Librarian: A Project Partner" published in 2019 describes the establishment and development of RDM and DMP support services at the University of Lille, and more specifically at its library. Since 2017, when they were already supporting researchers in some RDM tasks, they have been developing and adding new services to date. They divide their support into three main aspects: support for grant proposal writing, monitoring of open access compliance and support for RDM activities.

Regarding grant writing, the approach was applied to the different projects to analyze researchers' needs. In addition, the approach helped to build collaboration between researchers and institutional parties involved in RDM (e.g., the Grants Office, DPO, IT services), external consultants and experts in open science and RDM.

Regarding the monitoring of open access compliance support, authors show that librarians control all of the project's publications according to the Open Access

82 https://www.ru.nl/research-information-services/
83 https://dans.knaw.nl/nl/
84 https://repository.ubn.ru.nl/
85 https://dmponline.dcc.ac.uk/

policies; support researchers with licensing and access issues; improve metadata and review publications. Moreover, they collaborate with the repository manager, suggesting some improvements or changes if needed, guaranteeing the quality of the publications.

Finally, regarding the third aspect of RDM support for researchers, the ability to accompany researchers in writing a grant proposal is important because it gives librarians and researchers more time and opportunity to create a high-quality and detailed DMP, establish data management rules, and build trusting relationships. This collaborative approach also helps develop an approach to DMP support and identify important aspects of improvement.

To sum up, similar to our work described above and labeled in this chapter as "Five Case Studies" [145], their support is based on different activities. These activities include interviews with researchers to understand the context of the project, clarify researchers' issues related to data type, licenses, repository choice (e.g., institutional repository LillOA[86]), metadata, additional documents for confidential and personal data, and lead to the creation of the first draft of the DMP for the following validation by researchers. In other words, they accompany the researchers during the project, helping them at the beginning with the proposal and then with the creation of the DMP, improving their data management during and after the project, avoiding difficulties, which might arise, providing advice and monitoring sessions if necessary.

With the development of these services, the University can also show that they support researchers in all RDM and Open Science obligations, providing them adequate tools, training sessions, and workshops and helping to clarify their issues when they arise, motivating and encouraging researchers to comply with requirements.

14. Policy needs [58]: This paper, published in 2019, describes the establishment of RDM services and policies at TU Delft[87]. The university staff, in collaboration with the Data Stewards as well as library staff realized the importance of developing and implementing RDM policies[88], and in 2015 began to pay more attention to this topic along with the development of RDM services. During this development, they established many contacts with different stakeholders in the university, involved researchers in RDM issues, analyzed the needs of researchers and turned university projects according to the requirements of RDM and Open Science.

Along with the policy development, they started to develop RDM infrastructure, their technical part for supporting researchers in all RDM aspects and information knowledge part, organizing training sessions and workshops for staff, students, and researchers. They also launched the Data Stewardship program, which helps to improve the knowledge of the staff pertaining to RDM activities. They dedicated data steward at every faculty to provide more quality disciplinary support depending on the specifics of each domain. The data stewards in this context were seen as "first-line support", where researchers could clarify any issues related to the RDM. Moreover, data stewards constantly analyze researchers' difficulties and needs through surveys, individual meetings and consultations, and then suggest improvements for RDM infrastructure at the University and RDM policy.

Currently, the university has 4TU.Centre for Research Data[89], Service Desk[90] and Self-service portal where researchers and students can request support[91], Data stewards support[92], certified and trusted repositories for publications, data and heritage[93], DMP platform based on the DMP Online tool[94], and online courses and

86 https://lilloa.univ-lille.fr/
87 https://www.tudelft.nl/
88 https://www.tudelft.nl/en/library/research-data-management/r/policies
89 https://data.4tu.nl/info/en/
90 https://www.tudelft.nl/en/student/ict/service-desk
91 https://tudelft.topdesk.net/tas/public/login/saml
92 https://www.tudelft.nl/en/library/research-data-management/r/support/data-stewardship
93 https://repository.tudelft.nl/
94 https://dmponline.tudelft.nl/?perform_check=false

training sessions for researchers and support staff[95]. All of these efforts prove the university's motivation for the topic of RDM and show hard work to build a robust RDM infrastructure, underscoring the university's vision of open science policy and support for researchers.

15. Miami [9]: The University of Miami[96] in 2016 created a report and recommendations related to RDM infrastructure development. This report is a technical internal report that has never been published; however, it includes several important aspects that could be useful for other institutions in their development. Thus, Research Data Service, Libraries staff, Information Technology, the Office of Research and the Center for Computational Science at the University collaborated together aiming to implement RDM infrastructure and develop RDM services in accordance with the Open Science policies and funder requirements. Since 2008, they have focused on the creation of the data repository and DMP services, including researchers' consultations in all RDM aspects, on the improvement of the RDM curricula for researchers, students and staff, the creation of workshops and courses, and on the analysis of the universities needs according to the Open Science policies. Their work focused on collaboration with researchers and students, complimented by formal and informal meetings, collecting data from surveys and from workshops. Moreover, they establish the collaboration with the other institutions, such as the Calder Library at the Miller School of Medicine[97], the Rosenstiel School of Marine and Atmospheric Sciences (RSMAS)[98] or the School of Nursing and Health Studies[99] to analyze their way of developing RDM infrastructure and to share their own experiences.

Several pilot projects also helped in the development of the University RDM policies, namely in license issues, data deposit in the institutional repository[100], provision DOI for datasets through the University of California Digital Library's EZID service[101], subscription with Archivo Digital de Prensa from Columbia[102], and development of the RDM curricula together with the RSMAS and the Richter Library[103].

Currently, the RDM institutional policies are being constantly monitored and improved by collaboration with the Office of Research, UM Innovation, Libraries and university staff[104]. They also develop different guidelines, such as the Data Sharing Guideline, documents, such as Handbook, and templates for DMP[105].

Summing up, the author of this report highlights the importance of the establishment of the collaboration with different departments and stakeholders inside and outside of the university, developing and implementing support for researchers both in technological and information aspects, creating clear policies and guidelines, provide adequate tools for researchers with aiming to develop, implement, monitor, improve, and maintain RDM policies, infrastructure, and researchers/students support according to the RDM, funder and Open Science requirements. For DMP support services author proposes the following improvements: give more attention to the support of the data repository infrastructure, improve assessment and consultation services, expand support related to the DMP writing, reviewing, and submission, monitoring and analyzing provided tools, develop new ones, such as visualization tool, follow RDM, funder, and Open Science requirements and practices presented by others in the community, and constantly train the staff.

95 https://www.tudelft.nl/en/library/research-data-management/r/training-events
96 https://welcome.miami.edu/
97 https://calder.med.miami.edu/
98 https://www.rsmas.miami.edu/
99 https://www.sonhs.miami.edu/
100 https://www.library.miami.edu/data-services/finding-data.html
101 https://ezid.cdlib.org/
102 http://onbase.cinep.org.co/AppNet/
103 https://www.library.miami.edu/research/workshops-tutorials.html
104 https://www.it.miami.edu/about-umit/policies-and-procedures/index.html
105 https://www.library.miami.edu/research/workshops-tutorials.html

16. Learn from a Swiss [226]: A paper published in 2018 describes a collaboration between the École Polytechnique Fédérale de Lausanne[106] and ETH Zurich[107] to improve RDM services and DMP support. They analyzed researchers' needs and evaluated the human and technical resources that had already existed at the university to improve the RDM infrastructure according to the funder requirements and to improve DMP support.

Moreover, the Swiss National Science Foundation[108] - is one of their principal funders, started to require a specific DMP form for a grant proposal, so the university librarians offered a DMP template that included suggestions and recommendations and helped researchers create their plans more easily. This template is a "ready-to-use" document based on the funder requirements and adapted for researchers, reducing the load on librarians to support the DMP creation process. In addition, they developed various guidelines, training sessions, and workshops to familiarize and engage researchers with all RDM aspects.

The DMP template was created by a working group that included librarians, Vice President for Research and the Vice President for Information Systems, responsible for Research Ethics, and staff from the Technology Transfer Office, and the Research Office. The collaboration demonstrates recognition of the importance of DMP support needs and the motivation of university staff and stakeholders in establishing a well-organized RDM infrastructure. With this DMP template, researchers have easier and quicker answers to all requested questions. Moreover, suggestions of the answers and examples of plans allow them to understand what type of information should be on the plan, motivate researchers and make them more confident in RDM issues.

After the implementation of the DMP template, librarians organized a live demo of its use in the workshop, answering all questions of researchers.

The other DMP support that librarians provide is feedback after the reading and analysis of the plan. They give feedback during individual meetings, trying to clarify all issues and advise on adequate tools. These actions increased researchers' interest and made the University Library the first point for RDM and DMP support services.

17. Minnesota [156]: In 2019, the librarians from the University of Minnesota[109] created a strategic plan related to the establishment of improvement of the RDM infrastructure. This plan is an internal report and has never been published, and does not focus on the current situation at the University. However, it shows topics which should be implemented to make the University conform to the RDM and funder requirements.

Firstly, they focus on the expansion of the RDM and DMP support with a "rent a data manager" service that can help researchers during the creation of the DMP, its monitoring and reviewing or other RDM tasks. The RDM support, in this case, should be directed to the old, new, and emerging needs of both researchers and funders, forcing staff to monitor and analyze the relevance of all RDM infrastructure and emerging innovations in this area constantly.

Improvement of the knowledge through training sessions, workshops, and other courses, such as "Data management Bootcamp"[110], is another important point that requires university effort. Moreover, training sessions should be elaborated not only for researchers but also for staff.

FAIRness, which includes findability, accessibility, interoperability and reusability, is the next point where it is necessary to implement improvements. The authors of this report suggest establishing institutional policies and procedures and implementing long-term data stewardship.

106 https://www.epfl.ch/en/
107 https://ethz.ch/en.html
108 https://www.snf.ch/en
109 https://twin-cities.umn.edu/
110 https://libguides.umn.edu/datamanagement/dmbootcamp

To the moment, the University of Minnesota has already implemented a data repository[111] with an established policy of the collection, human participant, end-user access and preservation[112], DMP support in the creation of the first version of the plan, that researchers can require by contact form[113] or by email, metadata consultation and training sessions. Librarian staff also has 24/7 chat with real people without bots who can clarify issues related to the RDM[114].

To achieve the goals indicated in the strategic plan, librarians, together with other stakeholders of the university, organized the Research Data Services Strategic Planning Task Force, where needs and services are analyzed, and a new collaboration for expansion of the potential is established, and recommendations are defined and suggested.

18. Current trends [217]: In 2019, N.S. Redkina, in their paper, described the necessity of the development and implementation of the RDM services and support, including DMP at the Siberian Branch of the Russian Academy of Sciences[115]. Analyzing the literature related to RDM, and Open Science policies, they recognized RDM issues as a key factor in the development of science. Funder requirements and FAIR principles were also analyzed. In this context, to provide RDM and DMP support for researchers, librarian staff created different guidelines and training sessions to help researchers get familiar with RDM aspects[116]. The guidelines cover most of the RDM aspects, such as data protection, data storage, copyright, and data citation, among others. They also provide researchers support by email, in person, in the library, or by On-line consultation form[117].

Based on the analysis of the different sources, templates, and tools that existed in the scientific community, the institutions developed a Russian-language version of the Research Data Management Manual[118] and DMP Template[119].

Although academia still does not have a well-defined RDM infrastructure, neither the technical nor the policy part, librarians support researchers in all RDM tasks, considering RDM as an important topic, which will enable to make the institution according to EU standards.

In brief, an analysis of these papers shows that institutions are motivated to develop or improve their RDM infrastructure according to the RDM, funders, and Open Science requirements and policies. They understand the importance of RDM and DMP support and provide various ways to assist researchers in all aspects related to RDM. Elaboration and implementation of the institutional RDM policies, development or using more adequate tools for RDM and templates for DMP, primary role in researchers supporting, improving researchers, students and staff knowledge related to the RDM issues, and internal and external collaboration are the common aspects that are described in each selected article or report. These aspects are considered to be the most important for both RDM and DMP support, and for the development of the RDM infrastructure in general. Moreover, they help to describe our findings from the point of view of different dimensions. In the next section, we describe our findings in more detail and use the discussion to answer the Research Questions for a Systematic Review posed in this chapter.

111 https://conservancy.umn.edu/handle/11299/166578
112 https://conservancy.umn.edu/pages/drum/policies/
113 https://conservancy.umn.edu/contact/
114 https://apps.lib.umn.edu/qwidget/index.html
115 https://www.sbras.ru/en/index
116 https://tinyurl.com/3tsc5ywp
117 http://helper.spsl.nsc.ru/chat?locale=ru
118 https://tinyurl.com/3276h67y
119 https://tinyurl.com/yc4853ra

Table 3: Dimensions of analysis of the selected papers and reports

N°	Label	Title	Aspects							
			DMP support	Policy	Tools	Roles involved in support initiate	Training staff	Training students/ researchers	Internal collaboration	External collaboration
1	YOUth	FAIR, safe and high-quality data: The data infrastructure and accessibility of the YOUth cohort study			+	Data managers; IT department; Library	+	+	+	+
2	HongKong	Research data stewardship at the University of Hong Kong	+	+	+	Office of Knowledge Exchange; Library; Research Service Office; Technology Services Support	+	+	+	+
3	KU Leuven	Research data management and the evolutions of scholarship: Policy, infrastructure and data literacy at KU leuven	+	+	+	Central RDM support desk; library staff; IT services; Legal Department; Research and Development Office; Research Coodrination Office	+	+	+	+
4	Small-scale KTH	Establishing RDM services for small-scale data producers at big universities (KTH Institute)			+	KTH libraries, archive staff, It staff, Research Office and the Sweden National Infrastructure Computing PDCentre		+	+	
5	Small-scale WWU	Establishing RDM services for small-scale data producers at big universities (WWU)		+	+	University libraries, IT services, and administration		+	+	+
6	ETH Zeurich	Who does what? - Research data management at ETH Zurich	+		+	Human Resources department, the ETH Library, the ETH Zurich IT Services and the Research Data Management and Digital Curation Group at ETH Library	+	+	+	+
7	ETH Zeurich2	Forschungsdatenmanagement an der ETH Zürich: Ansätze und Wirkung	+		+	Human Resources department, the ETH Library, the ETH Zurich IT Services and the Research Data Management and Digital Curation Group at ETH Library	+	+	+	+
8	Three-pillar	Expanding the research data management service portfolio at bielefeld university according to the three-pillar principle towards data FAIRness	+	+	+	Competence Research Data Centre, library and computer centre	+	+	+	+
9	Collaboration	Leading fair adoption across the institution: A collaboration between an academic library and a technology provider			+	Library, Compliance Officer, and the General Counsel Office, Chief Security, Privacy Officers, the Office of Research, IT staff and leaders from the Office of Research		+	+	+
10	Peking	Research data management implementation at Peking university library: Foster and promote open science and open data	+		+	Library, University Management Research Data Centre, and collaborated with the Office of Scientific Research and the Social Science Research Office	+	+	+	+
11	RDM planning challenges	Exploring research data management planning challenges in practice			+	Support RDMP		+		
12	Five Case Studies	Institutional Support for Data Management Plans: Five Case Studies	+		+	Data steward, DPO, IT staff			+	+
13	Radboud	The role of CRIS's in the research lifecycle. A case study on implementing a FAIR RDM policy at Radboud University, the Netherlands	+	+	+	Support service desk		+	+	+
14	Partner	The embedded research librarian: A project partner	+		+	Library, Grants Office, Data Protection Officer, IT services or external consultants		+	+	
15	Policy needs	Policy needs to go hand in hand with practice: The learning and listening approach to data management	+	+	+	Data Stewards, library staff, 4TU.Centre for Research Data, Service Desk	+	+	+	+
16	Miami	Data Curation Initiative University of Miami	+		+	Research Data Service, Libraries staff, Information Technology, the Office of Research and the Center for Computational Science	+	+	+	+
17	Learn from a Swiss	Keep Calm and Fill in Your DMP: Lessons Learnt from a Swiss DMP-Template Initiative	+		+	Librarians, Vice President for Research and the Vice President for Information Systems, responsible for Research Ethics, and staff from the Technology Transfer Office, and the Research Office.		+	+	+
18	Minnesota	Strategic Plan for Research Data Services at the University of Minnesota Libraries	+		+	Librarians, Research Data Services Strategis Planning Task Force group	+	+	+	
19	Current Trends	Current Trends in Research Data Management	+		+	Library				

4.3 FINDINGS AND DISCUSSIONS

The selected papers demonstrate that the referred institutions are trying to develop or improve their RDM infrastructure and that they are focused on several aspects that we defined as dimensions of analysis. To summarize this information, we created Table 3, in which we indicate which aspects each of the analyzed papers contained. The systematic review allowed us to indicate the existence of the well-organized and implemented DMP support at the institution, officially recognized and implemented RDM policy, set of the RDM tools, roles in RDM and DMP support, different training sessions for staff and researchers/students, as well as internal and external collaboration as most important aspects that each institution should focus on to be in accordance of the RDM and Open Science requirements. We present a combined analysis of the selected paper, using its label and the 8 dimensions: DMP support, policy, tools, roles involved in support initiate, training sessions for staff, training activities for researchers/students, and internal and external collaboration. Note: the paper labeled Small-scale will be divided into Small-scale KTH and Small-scale WWU because they describe the two institutions separately. So the total number of studies for analysis is **19**.

First of all, the analysis of the selected papers shows that all institutions pay a lot of attention to the issues related to the development of RDM infrastructure and DMP support for researchers and students. Some of them have already installed DMP support services, while others still have them under development. However, all institutions offer different ways of helping researchers with RDM tasks in one way or another. If the institute does not have a well-established DMP support service, there is some training sessions related to general RDM issues.

As it is seen in Table 4, **14** out of **19** following papers and reports help us answer the first research question defined in this systematic review: "What kind of DMP support do institutions provide for researchers?". These are HongKong [285], KU Leuven [275], ETH Zurich [260], ETH Zurich 2 [202], Peking [187], Five Case Studies [146], Radboud [123], Three-pillar [230], Partner [93], Policy needs [58], Miami [9], Learn from a Swiss [226], Minnesota [156], and Current Trends [217]. As a result of our analysis of the first dimension, we have compiled a list of common and principal ways in which institutions provide DMP support to researchers:

Table 4: Studies describing different ways of the DMP support at the institution

N°	Label	Studies describing different ways of the DMP support at the institution (Yes / No)
1	YOUth	No
2	HongKong	Yes
3	KU Leuven	Yes
4	Small-scale KTH	No
5	Small-scale WWU	No
6	ETH Zeurich	Yes
7	ETH Zeurich2	Yes
8	Three-pillar	Yes
9	Collaboration	No
10	Peking	Yes
11	RDM planning challenges	No
12	Five Case Studies	Yes
13	Radboud	Yes
14	Partner	Yes
15	Policy needs	Yes
16	Miami	Yes
17	Learn from a Swiss	Yes
18	Minnesota	Yes
19	Current Trends	Yes

– supporting the researcher by experts, librarians, data stewards, etc. in the creation of the DMP (clarifying issues related to the DMP and the RDM in general without the plan creation);

– creation of the first DMP project by experts, librarians, data stewards, etc;

– analyzing existing DMPs, providing feedback and suggestions for improvement without monitoring;

– analysis of the existing DMP, providing feedback and suggestions for improvement, followed by monitoring;

– inclusion of the RDM experts on the project team since the grant proposal writing;

– validation or approval of the first draft of the DMP by experts;

– guidance provision;

– organization of the training sessions, workshops, courses, seminars;

– creation of the on-line resources (e.g. videos, posters, schemes);

– consultations and advisory services.

The same papers and reports help us respond to the second research question defined in this systematic review: "How is this support developed, implemented and organized?":

– service desk;

– RDM support departments;

– first-line support on the library;

– requests for DMP support by e-mail, phone, forms on the website, on-line chats;

– requests of feedback or analysis of the existed DMP;

– ticketing systems;

– self-services portal;

– individual or group consultations;

– use of the DMP templates;

– use of the DMP tools and platforms;

– use of the DMP checklists or other guidelines;

– "RDM expert rent" from the beginning of the project.

Some of these actions are implemented in institutions by means of automatic notification, which RDM experts receive officially by email.

Table 5 shows that almost all of the analyzed institutions (**17** out of **19**) provide training sessions for researchers or students. This type of activity can be seen as researchers' support in RDM issues because it also helps researchers to get familiar with existing RDM requirements and tasks, including DMP creation, and facilitate their execution. Moreover, the results confirm that if the institution does not have any DMP support implemented as a service, there are various training sessions

Table 5: Studies related to the training sessions for researchers/students

Nº	Label	Studies related to the training sessions for researchers/students (Yes / No)
1	YOUth	Yes
2	HongKong	Yes
3	KU Leuven	Yes
4	Small-scale KTH	Yes
5	Small-scale WWU	Yes
6	ETH Zeurich	Yes
7	ETH Zeurich2	Yes
8	Three-pillar	Yes
9	Collaboration	Yes
10	Peking	Yes
11	RDM planning challenges	Yes
12	Five Case Studies	No
13	Radboud	Yes
14	Partner	Yes
15	Policy needs	Yes
16	Miami	Yes
17	Learn from a Swiss	Yes
18	Minnesota	Yes
19	Current Trends	No

Table 6: Studies describing the implemented RDM policy

Nº	Label	Studies describing the implemented RDM policy (Yes / No)
1	YOUth	No
2	HongKong	Yes
3	KU Leuven	Yes
4	Small-scale KTH	Yes
5	Small-scale WWU	No
6	ETH Zeurich	No
7	ETH Zeurich2	No
8	Three-pillar	Yes
9	Collaboration	No
10	Peking	No
11	RDM planning challenges	No
12	Five Case Studies	No
13	Radboud	Yes
14	Partner	No
15	Policy needs	Yes
16	Miami	No
17	Learn from a Swiss	No
18	Minnesota	No
19	Current Trends	No

available. For example, papers labelled YOUth, Small-scale (both KTH and WWU University), Collaboration, and RDM planning challenges do not describe DMP support as a service or system in their institution. However, describe different training sessions for researchers.

The second dimension of the analysis is focused on the development and implementation of the RDM policy at the institutions. This analysis also help to answer both of the research questions posed at the beginning, since most of the analyzed papers and reports indicated RDM institutional policy as the most important aspect to be defined in an institution and officially endorsed. In other words, the development of the RDM institutional policy helps to develop RDM and DMP support in a better way at the beginning. Although only **6** out of **19** studies described its creation and implementation, almost all point to its development and implementation in the nearest future (Table 6).

In general, the RDM's institutional policy should describe the main rules and regulations concerning the collection, deposit, storage, sharing, licensing, access, preservation, publication, re-use and ethical aspects, among others. In other words, the RDM institutional policy defines the rules to be followed by researchers, students, university staff and other stakeholders in RDM activities regarding research

Table 7: Studies describing RDM tools for researchers support

N°	Label	Studies describing RDM tools for researchers support (Yes / No)
1	YOUth	Yes
2	HongKong	Yes
3	KU Leuven	Yes
4	Small-scale KTH	Yes
5	Small-scale WWU	Yes
6	ETH Zeurich	Yes
7	ETH Zeurich2	Yes
8	Three-pillar	Yes
9	Collaboration	Yes
10	Peking	Yes
11	RDM planning challenges	Yes
12	Five Case Studies	Yes
13	Radboud	Yes
14	Partner	Yes
15	Policy needs	Yes
16	Miami	Yes
17	Learn from a Swiss	Yes
18	Minnesota	Yes
19	Current Trends	Yes

data. Paper labelled HongKong, for example, describes that the University of Hong Kong started to develop its RDM infrastructure after the institutional policy formation, recognizing its importance and value for the planning of well-organized workflows, tools, and services [285]. The same importance to the policy given the KU Leuven University, is identifying its creation as a principal task and constantly analyzing and improving [275]. The Bielefeld University defines development as the principal and first pillar for the construction of the powerful, high-quality and well-designed RDM system [230]. Therefore, a policy should be created at the very beginning of the development of the RDM infrastructure, and preferably by a group of all university stakeholders.

Some of the institutions created different policies for each RDM task separately. For example, the TU Delft has three principal policies: funder's policies, publishers' policies and general policies[120], which in turn include research data framework policy, and research software policy, among others[121]; the University of Minnesota has data collection policy, human participant data policy, end-user access policy, preservation policy, terms of use, and deposit license, among others[122].

The next dimension of the analysis relates to tools to support RDM and DMP. These can either be developed in-house, for example, different types of storage described on YOUth paper: for biological samples, for tabular data and for other types [287], or already available in a scientific community (e.g. using the DMP Online tool described in the paper labelled as Radboud) [123]. The analysis shows that all papers and reports described different tools to support researchers, namely, **19** out of **19** studies used or developed tools for RDM tasks (Table 7). The most common ones are research repositories, data storage software and DMP tools.

No RDM and DMP support can be developed and provided without experts, institutional staff and specialists. Therefore, the next dimension of analysis is related to the roles within the institution that were assigned and defined to organize the RDM services, which also answers the Research Questions of our systematic review. As we supposed, **19** out of **19** studies have different departments that play a principal role in RDM and DMP support.

While almost every article gives the library a leading role, we have been able to compile a list of other possible units that could also help with RDM services and infrastructure in general:

120 https://www.tudelft.nl/en/library/research-data-management/r/policies
121 https://www.tudelft.nl/en/library/research-data-management/r/policies/tu-delft-faculty-policies
122 https://conservancy.umn.edu/pages/drum/policies/

– librarians;
– data managers;
– data stewards;
– IT department;
– office of knowledge exchange;
– research service office;
– technology services support;
– RDM support desk;
– legal department;
– research and development office;
– research coordination office;
– archive staff;
– computing/computer centre;
– administration staff;
– human resource department;
– digital curation group;
– research data centre;
– compliance officer;
– general counsel office;
– security chief;
– privacy officers;
– data protection officer;
– grants office;
– external consultants;
– vice-president for research and information systems;
– responsible for research ethics;
– among others.

The variety of roles shows that institutions have started to pay attention to issues related to RDM, and are trying to support researchers in different RDM tasks, hiring specialists in various areas and domains and creating official positions and offices for support.

The "principal role in supporting" dimension is related to the other one, namely "Internal Collaboration" (Table 8), which in turn show that to ensure and develop well-organized RDM and DMP support, many different stakeholders in an institution should collaborate together, be proactive, motivated, and focused on one goal, making their institution in accordance with RDM and funders requirements, Open Science policies and FAIR principles, increasing institutions value and their organizational level not only before government evaluation but also for its researchers. Internal cooperation is an important aspect of building the RDM infrastructure in general, both for policy making and for the development of tools and training sessions, among others. For example, Five case studies describe the collaboration with the DPO and IT staff [146], the labeled paper Collaboration indicates partnership with many offices and departments such as Compliance Officer and the General Counsel Office [188], among others. In general, this dimension is closely related to the other one, namely the "Principal role in supporting", which proves that in developing RDM and DMP support, everything is interconnected, and each dimension plays an important role.

External collaboration is the next valuable dimension related to improving all technical and theoretical issues related to RDM. Collaboration between institutions, between different national and international organizations, contributes to being up-to-date on changes related to the RDM taking place in the scientific community, promoting developed services, tools, knowledge, courses, and workshops, to analyze others' work and practices, learning from others' mistakes and avoid them in their own work, and to improve knowledge of the topic. Table 9 shows that **14** out of **19** studies described external collaboration that gave them good experiences and

Table 8: Internal collaboration at the institutions

Nº	Label	Internal collaboration at the institutions (Yes / No)
1	YOUth	Yes
2	HongKong	Yes
3	KU Leuven	Yes
4	Small-scale KTH	Yes
5	Small-scale WWU	Yes
6	ETH Zeurich	Yes
7	ETH Zeurich2	Yes
8	Three-pillar	Yes
9	Collaboration	Yes
10	Peking	Yes
11	RDM planning challenges	No
12	Five Case Studies	Yes
13	Radboud	Yes
14	Partner	Yes
15	Policy needs	Yes
16	Miami	Yes
17	Learn from a Swiss	Yes
18	Minnesota	Yes
19	Current Trends	No

Table 9: External collaboration at the institutions

Nº	Label	External collaboration at the institutions (Yes / No)
1	YOUth	Yes
2	HongKong	Yes
3	KU Leuven	Yes
4	Small-scale KTH	No
5	Small-scale WWU	Yes
6	ETH Zeurich	Yes
7	ETH Zeurich2	Yes
8	Three-pillar	Yes
9	Collaboration	Yes
10	Peking	Yes
11	RDM planning challenges	No
12	Five Case Studies	Yes
13	Radboud	Yes
14	Partner	No
15	Policy needs	Yes
16	Miami	Yes
17	Learn from a Swiss	Yes
18	Minnesota	No
19	Current Trends	No

results. For example, the WWU University collaborated with eScience Center[123] that provides support services and resources [268]; Peking University collaborated with the Beijing State Data Resource System and the National Information Centre in collecting valuable data across China [187].

The last two dimensions of analysis are related to the training sessions and knowledge improvement of both staff and researchers/students. These dimensions are also connected to support issues and help to answer our research questions. They are directly related because more training gives a higher quality of the staff/researchers' knowledge, which in turn allows easier DMP creation or execution of other RDM tasks, which in turn DMP and RDM support are better and easier. In other words, staff and researchers start to speak "the same language".

123 https://www.uni-muenster.de/EScience/

Table 10: Studies describing training for staff

N°	Label	Studies describing training for staff (Yes / No)
1	YOUth	Yes
2	HongKong	Yes
3	KU Leuven	Yes
4	Small-scale KTH	No
5	Small-scale WWU	No
6	ETH Zeurich	Yes
7	ETH Zeurich2	Yes
8	Three-pillar	Yes
9	Collaboration	No
10	Peking	Yes
11	RDM planning challenges	No
12	Five Case Studies	No
13	Radboud	No
14	Partner	No
15	Policy needs	Yes
16	Miami	Yes
17	Learn from a Swiss	No
18	Minnesota	Yes
19	Current Trends	No

And while training sessions for researchers/students are mentioned in almost every article (**17** out of **19**) (Table 5) and are given more attention because they can be seen as a way to DMP and RDM support, namely as a service, training for staff are not yet very common in the institutions. Only **10** out of the **19** studies described the importance of staff development (Table 10).

Along with the defined dimensions for review we also identify interesting points during our systematic review, which have influenced the analysis in one way or another, both positively and negatively.

Starting with the negative, many published papers describe general RDM issues. The description is more theoretical. Much attention is given to an introduction to RDM, funder requirements, Open Science policies and FAIR principles. Many studies also discuss researchers' needs and the necessity of the development of the RDM infrastructure and support services and systems in the near future. There are papers describing tools or templates that are used by institutions to support their researchers. However, there are few studies describing more specific technical and processual points, such as how to design and implement the services in an institution, which processes, workflows and stakeholders have been touched, analyzed and created to implement support as a service or system in the institution, what effort and how much time is required to provide well-organized services and support.

As our analysis shows, out of a total of **864** (Scopus) and **12** (RDA group) studies, only **56** (Scopus) and **6** (RDA group) were selected for the full-text analysis, and only **14** (Scopus) and **4** (RDA group) were retained for the in-depth analysis (Figure 50). Although some recommendations from the main funders (e.g. Horizon Europe) are already known and the relevance is already recognized, the DMP support and systems are still under development and not much information is available to reuse.

This shows that:

– the topic of DMP support development is still fairly new, which makes it acute and important to describe and disseminate in an academic environment;

– when developing RDM services and describing them, authors of the studies may not focus on DMP support and describe the support as a whole, making it difficult to apply existing practices in the institution which wants to develop or improve their DMP support;

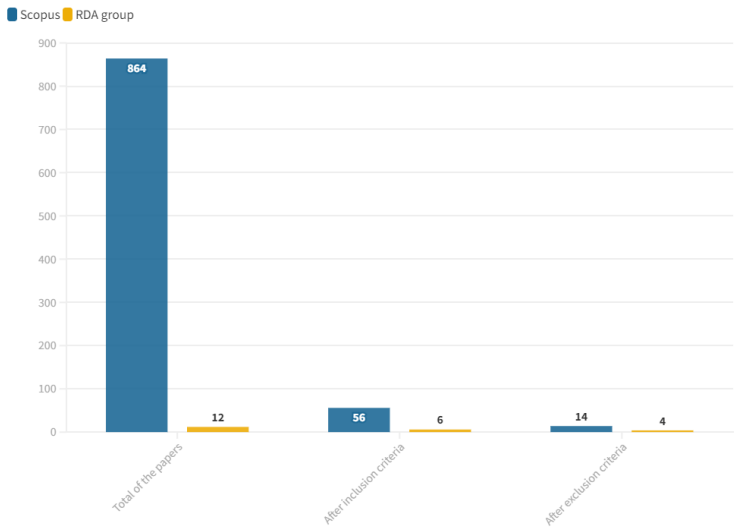

Figure 50: Studies related to the RDM issues and DMP support from Scopus and RDA group

– there is not-published technical and "grey" literature that may describe the issues of interest, but it is difficult to find and make use of it. This is exactly what the collaboration with the RDA group showed by sending its pointers to technical reports with comments such as *"Here is a report I wrote a while back that was never published outside of our institution ..."*, or*"You're indeed correct in assuming that there is a lot of grey literature on the topic floating around"*;

– there are some papers that are not in open access or some papers in another language such as German or French, that cannot be translated, which is also confirmed by the feedback from the RDA group: *´´In my experience, there is also – unfortunately – quite a substantial amount of publications done in German and French only, as the German and French-speaking countries are rather active in this area"*;

– there are papers describing the application of questionnaires or systematic reviews that only provide an insight into the stage of development of RDM at the institutions, and this type of publication is not suitable for our purposes.

On the positive side, we could point out that studies which do not cover in detail the design and implementation of DMP support and services in institutions anyway try to highlight the main and important points to pay more attention to when planning development. For example, some articles analyzing data repositories, their strengths and weaknesses, tools and platforms for implementing a DMP, putting the RDM infrastructure in line with FAIR principles or creating different protocols[124] which may also be useful during the support development process. This is also important to analyze because a well-designed DMP should contain detailed information about how the data will be collected, processed, stored, accessed and published. In other words, the creation of a support DMP should be a parallel process to the development and improvement of the overall RDM infrastructure of the institution. This can also allow increasing the value of the project, as the result of DMP support can be transmitted into a written, published and cited document.

Many studies also try to suggest their recommendations for the development of better services. For example: *"Establish clear policies and guidelines for research data governance and stewardship"*, *"Develop data management services including consultation on data management plans, policies, external research data repositories, and sensitive data"*, and *"Continued assessment and socialization of data curation practices"* [9].

Finally, we can state that the systematic review fully answered our Research Questions identified for this study and showed how much effort institutions should put into developing and implementing an RDM infrastructure that would be adequate

124 https://www.uu.nl/sites/default/files/dataaccessprotocol_youth_191029.pdf

for all needs, government, funders, and researchers and institutions' stakeholders. We also realized that development focuses on different dimensions and that all stakeholders should be connected with the same goal. Some institutions create collaboration with different departments within and outside the institution from the very beginning of the institutional RDM policy planning. Others start cooperating later during the implementation of different RDM tools, support systems and services when the need arises. For example, a connection with the DPO team can be established when the data steward starts supporting projects with sensitive and personal data. When a policy is still under development and not formally recognized, in most cases, support comes from librarians.

The systematic review also shows how much has already been developed in RDM-related areas, and that institutions and university staff should be constantly attentive to everything that is created, planned, tested and proposed for use. Only, by following and being aware of current information and applying the best practices of this community, is it possible to guarantee the quality of services provided and the level of the institution.

4.4 SUMMARY

Although there are not many published works describing the development and implementation of DMP support systems, the systematic review carried out in this chapter shows that the institutions try to provide different kinds of support to researchers in DMP creation and RDM issues in general. According to the systematic review, there are different ways to support researchers in the creation of DMP, starting with the development of the services and systems until the training sessions and creation of workshops and seminars.

In the context of this institutional support, researchers can receive help with issues related to metadata, data standards, ethical issues, and data repositories, which ultimately helps them to answer different questions when creating DMPs. For example, researchers can obtain lists of data repositories with an indication of their certification status[125] [67] or their grouping by scientific domains. Moreover, researchers can be informed about licenses, guidelines, best practices, and funding requirements, obtain templates for creating DMPs, and also request help to review DMPs.

According to the Goben, A., Zilinski L. and Briney, K. (2016) [102] most often, researchers look for support in creating plans from institutional librarians and our systematic review confirm it. Besides, some of the papers were published by university librarian staff and libraries were identified as one of the principal points for researchers' support. They are also not far behind in their RDM development and, now they provide support in RDM issues [259, 92, 3, 75] for researchers. Although libraries are still in the early stages of connecting to RDM infrastructure and need staff with RDM skills, we can see an expansion of their traditional mission. Some libraries already provide consultation services and include RDM support as new internal working groups [92, 102]. Others can organize meetings, e.g., in the offices or labs of researchers, receive researchers in the library, provide feedback on draft DMP documents by email, and collaborate with departmental grant administrators and project managers [92].

IT departments at institutions and research centres are another source of support for researchers in RDM activities in general, including creating DMPs. Often, partnerships between libraries and IT teams are established, and their goal is to develop new services to inform, train, and support researchers. In this context, and in addition to training programs, qualified personnel is hired, and efforts are made to ensure that institutions comply with RDM requirements [92, 3, 233, 75]. To cite but

125 https://www.coretrustseal.org/

a few, the University of Melbourne has a digital scholarship program [166], Cornell University implemented the Research Data Management Service [163], and the University of Glasgow [105, 167] developed a system for contacting researchers with approved projects and controlling requests for RDM support with automatically generated emails if a DMP is required, and the University of Sydney organized the eResearchUnit[126], which sends to researchers a pre-filled DMP template based on the abstract of the funded grant application.

The analysis of the literature shows that the existence of different tools and workflows to support researchers in RDM activities, including creating DMPs, aims to facilitate the daily work of researchers, simplify processes, and decrease the time spent creating DMPs while improving the quality of results. The work developed in this area shows that the diversity of scientific domains and respective plans requires people in charge of RDM support who are able to help with a multitude of requirements and adequate tools [167].

126 https://eresearch.uts.edu.au/

5 COLLABORATIVE DMP-BUILDING METHOD

Along with the different ways of the institutions' DMP support services described in the previous chapter, we propose the institutional collaborative DMP-building method that involves researchers, data stewards and other parties if required. This collaborative method was applied as part of an RDM workflow in research groups across several scientific domains. It has been proposed in accordance with existing RDM and FAIR requirements and best practices related to DMP support system described in Chapter 3. Moreover, this collaborative method is a basis or training for the DMP support system.

In this context, we describe the application of the collaborative DMP-building method and illustrate it through a set of case studies of projects from different scientific domains at INESC TEC. We also address the DMP monitoring process during the life cycle of projects and describe preliminary results obtained during collaboration with researchers, including their feedback.

In this chapter, we also describe the application of this collaborative method on projects from the Psychology scientific domain in order to define controlled vocabularies for DMP, which, in turn, help to create Domain Data Protocols. We conclude with the proposal of the Domain Data Protocols for the Psychology scientific domain that can simplify and standardize the DMP creation process. We also underline that there is motivation to improve the DMP support process according to the machine-actionable DMPs concept and to the best practices in each scientific community.

This chapter is based on the published paper Karimova, Y., et al. "Institutional support for data management plans: five case studies" 2020 [146] and its extended version Karimova, Y., et al. "Institutional support for data management plans: case studies for a systematic approach", 2021 [145].

5.1 COLLABORATION BETWEEN DATA STEWARDS AND RESEARCHERS IN DMP

In the context of the TAIL project, at the INESC TEC research institution, an RDM workflow was proposed. It included a set of RDM tools used to help researchers with tasks related to their data in the project (Chapter 2, Section 2.7), covered important stages of the data lifecycle [135] and considered the needs of researchers and the requirements of INESC TEC and funders. The set of RDM tools illustrated in Figure 51 supported researchers in different RDM activities and included the DMPOnline tool[1] for plan creation, the LabTablet[2] tool for data collection, the Dendro platform [3] for data organization and description, and the research data repository of INESC TEC [4] [136, 135] for data publishing.

As part of an institutional commitment to Open Data in research, the current RDM workflow and tools presented in Figure 51 needed to be refined according to RDM and funder requirements regarding DMP [259, 98]. Moreover, our constant collaboration with researchers has also shown that they are beginning to feel the need to provide DMPs and to seek support for their creation. In this context,

1 https://dmponline.dcc.ac.uk/
2 https://github.com/feup-infolab/labtablet
3 https://github.com/feup-infolab/dendro
4 https://rdm.inesctec.pt/

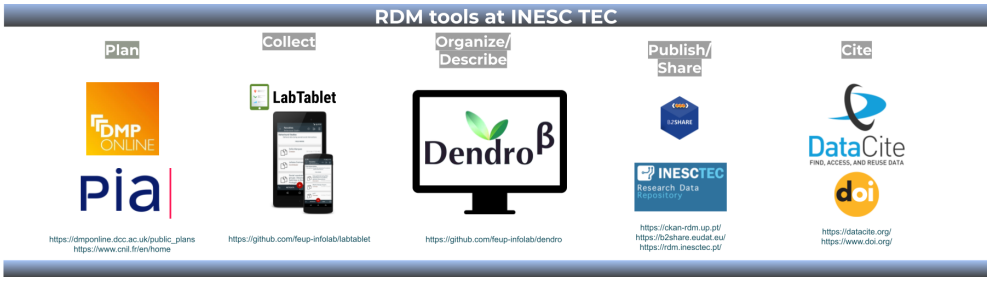

Figure 51: The set of RDM tools that was used at INESC TEC

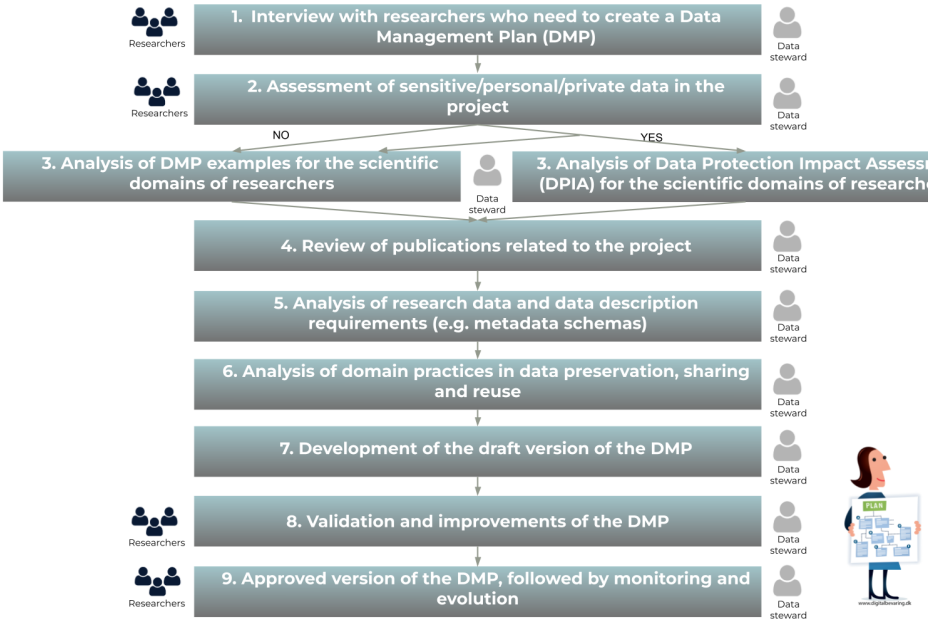

Figure 52: The collaborative DMP-building method

the elaboration of the collaborative DMP-building method started under the TAIL project, according to the requirements of the researchers and their groups.

Because funding agencies require a DMP submission alongside grant applications, we proposed a collaborative method where data stewards and researchers collaborate to prepare DMPs [144]. This collaborative method has been tested in the RDM workflow at INESC TEC with researchers from different scientific domains and includes the sequence of activities shown in Figure 52.

Firstly, the data steward interviews the researchers to comprehend their project, understand the context, and define what data will be collected and how they will be managed (Step 1 in Figure 52). Moreover, in this first meeting, the data steward can identify the difficulties of the researchers and evaluate their knowledge about RDM issues. Researchers face many difficulties when creating DMPs. For example, they may not understand how to organize data and may try to deposit all the data processed during a project into a single archive file. Another common issue is the existence of sensitive, private, or personal data. In one of the cases, researchers were unaware that the data collected in interviews involved personal data that incurs specific management requirements. Therefore, the first meeting with researchers has an important role, and it is the starting point of the data steward's support. This interview is also useful to collect information about publications and published data related to the project if any. That helps to identify the appropriate data repositories and metadata standards. Researchers are often unaware of any repositories or metadata standards for their domain.

After that, the data steward determines if any sensitive, private, or personal data may be collected during the project (Figure 52, Step 2). This kind of data often requires additional analyses and documents. For example, sometimes researchers need to get authorization from the ethics committee concerning research related to medical experimentation and human subjects. Moreover, researchers may need to contact a DPO, who helps to ensure that their project complies with any applicable data protection rules, such as the GDPR [201], and helps to create a Data Protection Impact Assessment (DPIA) that can minimize data protection risks. Other documents, such as confidentiality or intellectual property agreements may be required. In this case, the data steward helps researchers to: identify that such documents need to be created; establish a connection with the DPO; analyze the correctness of the informed consent, where interviewees give permission for participation; analyze the DMP examples; and, if a DPIA is required, analyze the DPIA examples for the corresponding domain (Figure 52, Step 3). These activities help the data steward to deepen their understanding of the project and prevent unpleasant situations related to the data, e.g., the lack of the informed consent signature or authorization from the ethics committee can be corrected at the beginning of the project.

In the next step of the collaborative method (Figure 52, Step 4), the data steward analyzes all the researchers' publications related to the project. This helps the data steward to find detailed information about the methodology, software, types and names of instruments that can be used in the project. It also helps to determine which documents are relevant, such as data reuse agreements that need to be preserved or published alongside the datasets. The next step is an in-depth analysis of the project research data and the corresponding description requirements (Figure 52, Step 5). Then, the domain practices in data preservation, sharing, and reuse should be identified and assessed (Figure 52, Step 6). This allows the data steward to propose an appropriate metadata scheme, estimate the amount of space required in the repository, and identify which file types and formats are most commonly used in the project.

With the necessary information collected, the data steward creates the first version of the DMP (Figure 52, Step 7). In some cases, interaction with the ethics committee is promoted. Furthermore, in case a DPIA is required, interaction with the DPO is proposed. In this case, the data steward helps to connect the researchers with the DPO and monitor their collaboration.

In the next step (Figure 52, Step 8), a draft of the DMP is presented to the researchers for validation and improvement. This interaction might take place over email. However, we recommend organizing a meeting with representatives from all project partners to clarify the authorship and ownership of the data and possible embargo periods, which may require interactions between project partners. The DMP (and the DPIA, if there is one) is not public at this stage. It will become public after all project partners authorize the publication of the approved version of the DMP (Figure 52, Step 9). The project leaders and the data steward will decide where the DMP will be published, e.g., through DMPOnline or Zenodo.

After the publication, the DMP is added to the project as a formal document for further monitoring. The data steward helps researchers understand the importance of keeping the DMP synchronized with any changes occurring in the project, i.e., regarding it as a "living" document. In a best-case scenario, the researchers themselves contact the data steward to announce changes or ask to create a monitoring schedule according to the project's duration. Nonetheless, the data steward reminds researchers about updating the DMP, e.g. by sending an email suggesting a date for a meeting. The frequency of the monitoring process depends on the duration of the project, when the DMP was created, and the dimension of the project. For example, if the project involves many partners from different countries, and the project's direction changes frequently, we recommend a DMP monitoring every three months. If the project is well-established and has a duration of more than one year, we recommend a DMP monitoring every six months. For this process,

the data steward analyzes the changes related to RDM issues in the corresponding domains and, during the meeting with the researchers, analyzes the changes in the project and proposes corrections. After that, they create another version of the DMP and schedule the next DMP monitoring session. If the DMP is ultimately not created, or the project is not approved, the data steward still keeps in touch with the PI to simplify the creation of the plan for the next project, taking advantage of the knowledge obtained. This collaboration can lead to the joint organization of workshops, participation in hackathons, and contributions in papers. In the collaborative method described above, all processes related to creating and monitoring DMPs happen manually.

This collaborative DMP-building method was applied as part of an RDM workflow in research groups across several scientific domains with the goal of establishing a systematic approach. The set of case studies provides experience of this application of the collaborative method to analyze its value.

5.2 CASE STUDIES OF DMP WITH THE COLLABORATIVE METHOD

INESC TEC is a research institute with more than 700 researchers and 300 PhDs from different scientific areas such as computer science, media and communications, power systems, energy, robotics, industrial engineering, and bioengineering. For more than 5 years, INESC TEC has been nurturing experimental activities in RDM, partly as a research endeavour, having in mind the development of new services to support RDM in the context of running projects and to publish datasets in the institutional data repository. As part of an institutional commitment to Open Data in research, INESC TEC provides data stewardship as an experimental service. There is an established workflow for supporting researchers in organizing, describing, and depositing data. The strategy to create DMPs is part of this commitment and currently involves 2 part-time employees (one data steward and one repository manager) who promote awareness of RDM and process requests for DMPs.

The cases described next followed the collaborative method mentioned above, and involved people from several research groups, both at INESC TEC and outside. In each case, DMP creation started with a request sent by a researcher to the data steward. Sometimes researchers knew exactly where to send the request, but there were cases when researchers spoke with those who had already created plans to verify where to obtain help. Then they sent requests by email or requested in person. We used a collaborative method for DMP creation to help researchers. This method was the first approach to the establishment of the institutional support for DMPs and had to be tested before it was systematically applied.

Application of the collaborative method in real cases and different scientific domains allowed us to identify specific requirements and points of improvement. DMPs are typically created at the beginning of projects, but sometimes in the middle of the work, and rarely - at the end. We have worked on ten plans for projects related to Environmental radioactivity, Biodiversity, Education, Oceanography, Psychology, Environmental engineering, Health, Statistics, Biomedicine, and Artificial intelligence. Some of the plans were completed and published, while for others the preparation and monitoring continued afterwards. Some of the plans were created at the beginning of the projects, or even during grant proposal preparations, while others were created in the middle or in the final months of the projects. Some of the projects had been approved and some were waiting for a decision regarding the submitted proposals. Next, we will detail the DMP creation process for these ten case studies, highlighting the aspects that can be transferred to other cases. These cases were in different preparation stages, had different budgets, and gave us substantial information for analysis. In Table 11, we present the scientific domain, the names

Table 11: Summary of the DMPs created with the support of INESC TEC

Project domain	Project title	Contact person	Budget (K euros)
Environmental radioactivity	Gamma radiation monitoring campaign at the Azores ENA-ARM station (Graciosa Island)	Susana Barbosa INESC TEC, CS, NIS	No funding
Biodiversity	FARSYD: FARming Systems as tools to support policies for effective conservation and management of high natural value farmlanDs	Ângela Lomba UP, CIBIO, InBIO	≈ 181
Education	SCReLProg: Self and co-regulation in e-Learning of computer programming	Daniela Pedrosa UTAD, CIDTFF	≈ 230
Oceanography	SAIL: Space-Atmosphere-Ocean Interactions	Susana Barbosa INESC TEC, CS, NIS	≈ 100
Psychology	Identification of learning and development conditions at/through work: challenge the paradoxes of technological introduction and lifelong learning	Marta Santos, FPCEUP Mafalda Lopes, FPCEUP	Not approved
Environmental engineering	Future Cities: UrbanSense	Ana Aguiar FEUP, DEEC, CECF	≈ 1 615
Health	aMILE: application of text mining to clinical reports of patients with Acute Myeloid Leukemia	Rita Rb-Silva IPO-Porto	Proposal
Statistics, Health	COnVIDa. Mainstreaming well-being and quality of life in IPC measures for South European Nursing Homes	Alexandra Lopes FLUP, DS, DEI, DEGI	Proposal
Biomedicine	Facilitated Genome Editing as Responsible Research and Innovation	Pedro Ramos i3S	-
Artificial intelligence	TRUST-AI: Transparent, Reliable and Unbiased Smart Tool for Artificial intelligence	Gonçalo Reis Figueira INESC TEC, ISE	≈ 3 996

Table 12: Summary of the DMPs created with the support of INESC TEC (cont.)

Project title	DMP status	DMP creation period	DMP monitoring
Gamma radiation monitoring campaign at the Azores ENA-ARM station (Graciosa Island)	Published	At the end of the project	No monitoring performed
FARSYD: FARming Systems as tools to support policies for effective conservation and management of high natural value farmlanDs	Published	In the middle of the project	August 2020
SCReLProg: Self and co-regulation in e-Learning of computer programming	Published	At the beginning of the project	March 2020 August 2020 November 2020
SAIL: Space-Atmosphere-Ocean Interactions	Published	At the beginning of the project	April 2021
Identification of learning and development conditions at/through work: challenge the paradoxes of technological introduction and lifelong learning	No funding, DMP not published	During proposal evaluation for funding	No monitoring performed
Future Cities: UrbanSense	In revision	After project completion	No monitoring performed
aMILE: application of text mining to clinical reports of patients with Acute Myeloid Leukemia	Published	During proposal evaluation for funding	June 2021
COnVIDa. Mainstreaming well-being and quality of life in IPC measures for South European Nursing Homes	In preparation	During proposal submission	Waiting for funding
Facilitated Genome Editing as Responsible Research and Innovation	In revision	During the project	Will be scheduled after DMP publication
TRUST-AI: Transparent, Reliable and Unbiased Smart Tool for Artificial intelligence	In revision	At the beginning of the project	Will be scheduled after DMP publication

of the projects and DMP, the contact persons for the DMP creation, and the budget of the project for each of these cases. In Table 12, we present the status (published, not published, in revision, or in preparation), the project phase of its creation, and the scheduled monitoring sessions for each of the DMPs.

In general, during the creation of the plans, the data steward collaborated with the PIs of the projects, despite the fact that most of the projects had several partners and, consequently, many participants. Large-scale projects usually have different people responsible for the different RDM activities, e.g., one person for data collection and processing, and another for preservation and backup. Smaller projects usually have only one person responsible for all RDM activities, who may or may not be the PI. After the collaboration the researchers completed a questionnaire to measure their attitude towards DMP support and RDM issues in general [5].

5.2.1 Environmental radioactivity

The goal of the "Gamma Radiation Monitoring" project[6], conducted by INESC TEC, was to study the concentration of the noble gas radon (Rn-222). The aim was to examine how radon is influenced by meteorological conditions, how it impacts the local atmospheric electric field, and its association with the atmosphere's ionization and aerosol concentration. The PI of this project already had experience in RDM

5 https://forms.gle/zZMVVbRp9z77XXBA8
6 https://www.arm.gov/research/campaigns/ena2015grm

activities, but not in creating DMPs. The data steward held an interview with the PI to understand the project's context and collect data and papers related to it. After the interview, their analysis continued, identifying the absence of sensitive data, and studying specific requirements for data management and description in this domain. At the same time, several examples of plans in similar domains were analyzed, as well as some DMP templates, and the first version of the DMP was created. Moreover, the analysis of the plans' examples and DMP templates related to the same scientific domain helped to create a list with specific questions for verification and confirmation with the PI at the second meeting. Not all questions from the list were used in the DMP, however many of them helped to add more specific details about the tools used during the project and their calibration method, the software used, the data analysis, the measurement method, the data transfer process from the station, and even what should happen to the project data in case the PI left. The list showed that during the DMP creation, researchers prefer to have some suggestions for their plan, as they facilitate the creation, showing more options than they even remember. This DMP was created in the final phase of the project and did not require monitoring[7].

5.2.2 Biodiversity

The goal of the "FARSYD" project [158], conducted by the Research Center in Biodiversity and Genetic Resources (CIBIO), part of the Research Network in Biodiversity and Evolutionary Biology (InBIO) at the University of Porto, was to examine the relationship between farming systems, biodiversity, and ecosystem services in high nature value farmlands. The PI had experience in RDM but not in DMP creation. However, for this project, she created an Excel file with a detailed description of each experience in the project. This file and the description of the project helped the data steward to understand the project's context and identify the existence of private and sensitive data that cannot be publicly disclosed. Following the collaborative method, the data steward analyzed all the obtained information, collected examples of DMPs in the biodiversity domain, and experimented with the GFBio DMP Tool[8] and the Best Practice Guide [31] promoted by the German Federation for Biological Data. This analysis helped to prepare a list with specific questions that were validated with the PI. Two existing checklists were verified: the one prepared for the Environmental radioactivity project and the one specific to the Biodiversity domain. Although the first list of questions was not immediately applicable to this plan, some of the points were adapted and used. This led to the inclusion of the description of the specific tools used during the project, the software used, the training areas for habitat mapping, and several approaches used to obtain data depending on the specific target and location. This type of verification is not part of the proposed methodology; however, it is very important for our future work as it helps to identify differences between scientific domains during the DMP creation. The analyzed GFBio DMP Tool helped the researcher to add specific information such as Project type, with possible values Field Work, Observational, Simulation, Assimilation, Experimental, Laboratory, and Modelling, where Field Work and Observational values were considered more suitable for FARSYD. This tool also demonstrated to the researcher the existence of different metadata standards and legal requirements specific to the biodiversity domain. Examples of the latter include the IUCN Red List of Threatened Species and the Nagoya Protocol. Moreover, this project contained private data provided by the Instituto de Financiamento de Agricultura e Pescas (the national institute for funding agriculture and fisheries), the Integrated Administrative and Control System, and the Land Parcel Information Systems.

7 https://tinyurl.com/vusz7yca
8 https://www.gfbio.org/plan

The data steward helped to add detailed information on the management of this kind of data and the corresponding preservation rules with restrictions and different access levels. The first version of the DMP was created in the middle of the project after establishing the collaboration between INESC TEC and CIBIO-InBIO, but it was not published. By the end of the project, the plan was improved, detailed, and publicly shared through DMPOnline[9]. This DMP has been monitored and improved twice since the plan was created, with the last monitoring activity in August 2020. In this case, no new information was added or changed in the DMP and no new monitoring session was planned, given that the project ended shortly.

5.2.3 Education

The goal of the "SCReLProg" project[10], conducted by the University of Trás-os-Montes and Alto Douro (UTAD), the Universidade Aberta (UAb), the University of Aveiro (UA), the Research Center on Didactics and Technology in the Education of Trainers (CIDTFF), and INESC TEC, was to develop a pedagogical approach to overcome programming difficulties and effective strategies for self-and co-regulation of e-learning. The researchers in this project did not have experience with RDM. However, they followed the work of the TAIL project, and when it became necessary to create the DMP, they sent a request by email. Together with the support request, they sent a lot of documentation about the project which helped the data steward to understand the project's context before the first meeting. These documents were valuable to provide an overview of the project, prepare a list of questions for the researchers, identify the existence of personal data, and prepare information related to the informed consent, ethics committee approval, and conformance to the General Data Protection Regulation. All meetings were conducted via Skype with the researcher in charge of RDM. After that, the first version of the DMP was created and sent to the researchers for verification. The existing informed consent form created by the researchers was also improved. Moreover, the data steward created the first version of the DPIA because it was necessary to take proper care of personal data, which required validation and analysis by the DPO. In this context, the data steward sent an email with the draft of the DPIA to the DPO of INESC TEC and organized a meeting between the people in charge of the project, the DPO and his team. Due to the small risks and threats estimated by the DPO, the DPIA was considered optional and replaced with a detailed DMP and a signed agreement on data processing between the project partners. The published version of the DMP [11] was shared, corrected, and approved by the researchers in charge of the project. Six months after the start, the DMP was made public, and the first monitoring session was scheduled for March 2020. During this session, some points in the plan were changed. For example, some information about integrating new collaborators into the project was added, including their names and responsibilities, and new forms of data collection. The second monitoring session was scheduled for August 2020 and did not result in any changes. However, in November 2020, the researcher in charge of RDM emailed the data steward with a question related to sensitive data. They wanted to share all the collected data with all the collaborators of the project, which was against the original conditions defined in the DMP. This situation led the data steward to contact the DPO again, send the DPIA draft created at the beginning of the project, and review the necessity of a DPIA. Currently, a DPIA is considered necessary and is under revision. The data steward has been following this process, organizing meetings, and helping researchers improve the DPIA. The creation of a DPIA requires substantial time and effort, as it is necessary to respond to all specific questions related to personal data protection. To guarantee the anonymization of

9 https://tinyurl.com/2f4hw6kw
10 https://www.inesctec.pt/en/projects/screlprog
11 https://tinyurl.com/yeyb6u6x

the personal data, for instance, they needed more time to analyze the text of the students' answers, for which the project had no allocated resources.

Although the researchers are concerned with the creation of these new documents for their project, they appreciate the data steward's help. During the work with this case, the data steward suggested replacing the Google Drive storage with the institutional Drive at INESC TEC[12] and Google Forms with UESurvey, due to security issues. In this case, the DMP was created at the beginning of the project, after it was approved for funding. This helped to prevent complicated situations related to personal data before the data was collected, prepared, and shared while providing the project team with all the required documentation.

5.2.4 Oceanography

The DMP of the "SAIL" project[13], conducted by INESC TEC, Marinha (the Portuguese navy), the Atlantic International Research Center, the Interdisciplinary Center of Marine and Environmental Research (CIIMAR), and the University of Minho (UM), was the most complex and detailed plan created at INESC TEC due to the diversity of data, tools, software, and internal procedures associated with the project, and due to the number of institutions involved. Moreover, this was the first plan to comprise several scientific domains: biodiversity, oceanography, and robotics. The first version of the DMP was created faster than the plans of the other projects because the project was due to start, and the PI of the project, who had experience in RDM activities, wanted to define and validate all the important RDM issues before the start of the project. To create the first version of the DMP, the data steward followed the proposed collaborative method and sent the first draft version of the DMP to the PI for verification. The first DMP draft raised many questions related to ownership of the data. However, after 7 months, the first version of the DMP was published on Zenodo [15] with its own Digital Object Identifier (DOI). Zenodo also allows publishing subsequent versions of the DMP and citing this DMP as a publication. SAIL had created a community on Zenodo for the project, where all documents and outputs would be published[14]. A monitoring session was scheduled for April 2021, during which it was decided to add more information about new documents created during the project. The documents were created due to sensor reading errors that occurred during the project. They describe the process of correcting errors and creating the program that helped to do quality control. Moreover, they describe the post-processing procedure. Currently, this DMP is in the adjustment process. This case study shows the importance of creating DMPs at the beginning of projects in order to allow sufficient time to clarify any ownership issues before data collection starts, avoiding troublesome situations.

5.2.5 Psychology

The "Identification of learning and development conditions at/ through work: challenge the paradoxes of technological introduction, and lifelong learning" project, led by the Faculty of Psychology and Education Sciences at the University of Porto (FPCEUP), focused on the (re)learning of productive processes, the transformation of organizational models, and the improvement of the qualifications structure. This project involved several types of data and data collection techniques and was expected to deal with personal and sensitive data. This case study's specificity was the existence of a researcher in charge of RDM tasks, whose role was to plan, organize, and answer any RDM questions from the researchers. The person already had experience in data management, but not in creating DMPs. She contacted us to get the support of the data steward in creating the plan while improving her knowledge

12 https://drive.inesctec.pt/
13 https://tinyurl.com/ydgy6lng
14 https://zenodo.org/communities/sail/

of RDM. Moreover, she wanted to prepare and organize her project in advance, so the creation of the DMP started when the project was under evaluation for funding. In this case, DMP creation took an abbreviated path. The first version of the DMP was sent directly to the data steward for evaluation and correction. In two days, the data steward analyzed the plan, added comments, and identified the points that needed more detail. Moreover, the data steward also raised issues regarding informed consent and DPIA, the corresponding contact with the institutional DPO, and the ethics committee's approval due to the existence of sensitive data in the project. The first version of the DMP was prepared using the DMPOnline tool and maintained "closed" pending the approval of the project. Although the project was not approved, the person in charge of RDM confirmed that it was a positive experience that allowed her to acquire new skills. With the knowledge obtained, she could contribute to another project application using the available tools and taking into account the importance of DMPs. This experience, which was seen as a training session for creating DMPs, also gave her the tools to help other research groups that requested support and increased her confidence.

5.2.6 Environmental engineering

The DMP of the "Future Cities: Urbansense" project[15], conducted by the Municipality of Porto, Porto Digital, Instituto de Telecomunicações, the Center of Competence for Future Cities, UP, and the University of Aveiro, was started in 2019 in the context of an MSc dissertation in Information Science at the University of Porto. The "Future Cities" project focused on unleashing the full interdisciplinary potential of research in urban technologies and creating the UrbanSense platform, which, in turn, aimed to transform the city of Porto into a Smart City. The first version of the DMP was created in Portuguese and was not published. We knew that the PI had no experience with RDM issues but was interested in collaborating not only to create a DMP for this project but also for others. So, the PI was contacted to see if there was an intention to improve and publish the DMP. Although the project was finalized, the PI showed interest in publishing the plan. The data steward followed the collaborative method, analyzed all the information related to the project and the existing version of the DMP, created a draft, and sent it for validation. Moreover, the DMP was translated into English, which is more suitable for a large-scale project. Currently, the DMP is under revision, waiting for publication. The monitoring of the DMP was not scheduled due to the status of the project. This case study showed that DMPs could be created at different stages of projects, even after their end. Moreover, the collaboration with this researcher paved the way for a line of other projects that would also need DMPs.

5.2.7 Health

The "aMILE: Application of text mining to clinical reports of patients with acute myeloid leukemia" project, proposed at the Portuguese Institute of Oncology in Porto (IPO-Porto), is related to the creation and validation of text mining algorithms to extract relevant clinical data from PDF files (such as hospital discharge summaries and other medical reports) in a reliable, safe, and confidential way, transforming them into structured data. The PI of this project wanted to create the plan in advance and submit it with the proposal. Moreover, the PI was very interested in improving her knowledge of RDM issues and saw this collaboration as an opportunity for training. The DMP creation started with the first draft created by the PI, which was sent to the data steward. The plan took some comments and suggestions and was published on Zenodo with DOI assignment [239]. Although the project has not yet been approved for funding, the published DMP required some

15 https://cordis.europa.eu/project/id/316296/reporting/fr

changes, which were implemented at the monitoring session scheduled for the end of June 2021. This case study showed the engagement of the researcher in DMP creation, since before contacting us, she did a lot of previous work, looked for information, created a very comprehensive plan, contacted the ethics committee, verified the necessity of the informed consent, and showed great interest in the task.

5.2.8 Statistics, Health

The "COnVIDa. Mainstreaming well-being and quality of life in IPC measures for South European Nursing Homes" project, proposed at the Department of Sociology of the Faculty of Arts and Humanities of the University of Porto, was in the preparation stage when we started DMP creation. The PI, who had no experience with RDM, contacted the data steward to create the first version of the DMP in advance. The DMP was supposed to be created and submitted together with the funding proposal. Due to the urgency, lack of documents and other information related to the project, the data steward did not follow all the steps of the collaborative method. In this case, the interview provided minimal information for the creation of the first light version of the DMP, where only the main information related to the project was indicated, as well as the people in charge and the kind of data to be collected. However, this case study shows that the researchers started to be concern about the DMP at the early stages of their projects, namely when it comes to creating the DMP with minimal information to submit it along with the project proposal.

5.2.9 Biomedicine

The "Facilitated Genome Editing as Responsible Research and Innovation" project, conducted by the Laboratory of Animal Science of the Institute for Research and Innovation in Health (i3S) at the University of Porto, aimed to establish several tools and protocols to improve the Clustered Regularly Interspaced Short Palindromic Repeats technology related to efficient genome editing. Similarly to the "Future Cities: Urbansense" project, the first version of the DMP was created in the context of a master's thesis. It was also in Portuguese, not published, and not monitored. In this context, the data steward contacted the PI of the project to understand the project's state and the interest in the publication of the DMP. According to the PI, at the beginning of the project, the researchers had difficulties in understanding the "language" of the RDM requirements and in filling in the DMP due to their limited experience in RDM. Moreover, their project did not initially require the creation of a DMP. However, during the project, RDM and Open Science issues raised the need to review policies and GDPR questions. Since the creation of DMPs may require substantial time and effort, during the interview with the data steward the researchers wanted to understand the benefits of creating and publishing them. Moreover, they wanted to see examples of existing plans in their scientific domain and the plans published on Zenodo. The first meeting (by video conference) helped the PI of the project to understand the importance of improving, translating, and publishing the DMP, giving it a clear scientific value, and turning it into an output of the project. Thus, the data steward followed the collaborative method to create the DMP, starting with an analysis of the existing Portuguese version. The new version of the plan was translated and enriched with new information: the names of the people in charge of the RDM tasks; data on funding; additional documentation of the project that also will be preserved; detailed information about storage and preservation of the data. The kind of data collected during the project was better defined, as well as how it would be managed. The first version DMP is under revision by the PI of the project. When the revision is complete, the DMP will be published on Zenodo with the corresponding DOI and scheduled for a follow-up monitoring session.

5.2.10 Artificial intelligence

The request for support with creating a DMP for the "TRUST-AI: Transparent, Reliable and Unbiased Smart Tool for AI"[16] international project, conducted by INESC TEC, the University of Tartu, and the Leiden University Medical Center (LUMC), was sent to the data steward by email from an INESC TEC researcher. The project was approved for funding and related to the design of a transparent, reliable, and unbiased tool focusing on machine learning applications that use images, text data, and tabular data, including classification/regression. In this case, there are many entities with different responsibilities involved. For example, one of the responsibilities of INESC TEC is creating the DMP. However, the researchers had no experience with that. The specificity of this case is in the creation of the "Communication and dissemination" Working Package (WP), which aims to help promotion of the results to the scientific community. Moreover, in this WP, there is a task for "DMP creation". The data steward followed the collaborative DMP-building method and organized a meeting with the researchers. The data steward also collected all necessary documents prepared previously by the researchers, such as the project's proposal, detailed description of the use cases of the project, consortium and grant agreements, and related information for analysis, which helped to create the first version of the DMP. The document created earlier with the DPO of INESC TEC, which described the ethical issues and roles of each entity in the project was analyzed. It helped the data steward to understand the kind of data collected and how this data will be managed during the project. The collected information allowed the data steward to create quite a complete first version of the DMP, which is currently under revision. The data steward added many specific questions for researchers, which schould help to complete the DMP. To answer the questions, the researchers should organize several meetings with their project partners, since they cannot answer all the questions on their own. For example, they do not know who will be responsible for RDM activities in each use case of the project, which, in turn, is complicated by the number of countries involved. This is a case where the DMP is required for submission of the grant proposal and then constitutes an official document that will be monitored during the project, updated, approved, and followed by all project partners. The other specificity of this case is that the researchers have already scheduled monitoring sessions of the plan every 6 months after the publication of its first version. The preparation of the documents earlier in the project also shows the responsibility of the researchers and their understanding of the RDM requirements.

5.3 RESEARCHER FEEDBACK ON THE CASE STUDIES

There are many different ways for institutions to support researchers in DMPs. The collaborative DMP-building method described above was the first attempt to establish a DMP workflow for research projects and to ingrain RDM into project activities.

The DMP creation in the early stages of projects makes the lifecycle of the data explicit, gives enough time to prepare the necessary resources, and verifies ethical and legal issues, avoiding misunderstandings among project partners. This has a positive impact on institutional research maturity. Though a lot of simplification is sought when establishing institutional support, managing research data requires a significant level of insight into research processes and careful project preparation.

The case studies where the collaborative DMP-building method was applied show how important it is for institutions to provide DMP support and have data stewards who can clarify any questions that researchers might have as they develop their projects. The cases show the importance of having a DMP monitoring pro-

16 https://www.inesctec.pt/pt/projetos/trustai

cess and collaboration between data stewards and other parties as required, namely the DPO. Researchers might create DMPs based on existing examples from their domains on their own, which likely will not have an adequate level of detail and monitoring. Moreover, when creating DMPs without data stewards, researchers should clearly understand what information must be included. DMPs should not be confused with other documents required for projects, such as the Project Management Plans, which are also important and define how projects will be executed, monitored, and controlled [169]. Although Project Management Plans have much information about projects, such as goals, budget, timeline, and project partners, they do not focus on RDM issues [17]. Namely, they do not require information about what data will be collected, processed, shared, or reused. They also do not lead to compliance with the FAIR principles: the findability, accessibility, interoperability, and reusability of data produced in projects. However, having Project Management Plans can simplify the creation of DMPs and improve the understanding of projects in general.

The researchers of each case study, after collaborating with the data steward, responded to a questionnaire we had prepared in advance and provided interesting comments. The answers to the questionnaires[18] highlight the importance of data stewards. Some researchers declared that it was essential to have specialized staff to assist them in the tasks, and all stated that it was essential to have a data steward, who helped with RDM issues at the institution and helped to prevent errors during the planning stages. The researchers affirmed that they had become more confident about RDM issues, including creating DMPs. They had begun to understand the terminology better and had been acquainted with various repositories, tools, and applications that simplified the creation of DMPs and RDM in general. Finally, the researchers established contacts with other stakeholders and collaborators, such as the DPO, which can be useful for further collaborations. Even though some projects were not approved for funding, our collaboration in creating the DMPs allows the researchers to apply the accumulated experience and knowledge in new projects and to help others.

The cases described above were collected over a period of 4 years. The cases emerge from RDM, funder and FAIR requirements, as well as interest from researchers looking for RDM support and (more frequently) expecting to comply with a mandate by a funding body. The cases show that researchers have begun to pay more attention and value to DMPs, as well as that they have increased their engagement with RDM in general and commitment to DMPs in the planning stages of projects. They have started to see the creation of DMPs as an opportunity to define the entire data management strategy of projects at an early stage, which can influence the whole project. They have also started to contact the data steward in advance to avoid spending a lot of effort and time unnecessarily. The cases illustrate the application of the collaborative method in several scientific domains and projects of any budget dimension. Additionally, it can serve as a basis for establishing a DMP support service in a research institution large enough to justify an organized such a service.

Furthermore, the preliminary results show that the identification of sensitive and personal data is one of the main aspects where unexpected difficulties may arise. The case studies described above led to a better articulation between the data steward, the DPO, and the ethics committee, defining the corresponding responsibilities when creating documents for the project. The case studies show that projects with personal data usually demand collaboration with a DPO and in some cases the elaboration of a DPIA. That, in turn, requires prior knowledge about data processing steps. Thus, any work on a DMP should precede consideration of a DPIA. However, the recommendations from a DPIA may lead to a revision of the DMP, and therefore an iterative approach is suggested. The results of this work show that the DMP

17 https://www.simplilearn.com/what-is-a-project-management-plan-article
18 https://forms.gle/zZMVVbRp9z77XXBA8

monitoring process is essential because a DPIA can be required at any moment, due to changes in the project. An ethics committee deals with more fundamental issues like the appropriateness of a project's purpose and research methods. In many domains, approval by an ethics committee is mandatory, so it may already have been handled when a DMP is considered. In other cases, the timing of a project proposal allows the ethics analysis to be performed in parallel with the DMP process.

Metadata schemas, data organization and preservation rules, choosing repositories, and scheduling RDM tasks among project partners are other complex issues where researchers need support. To describe all DMP elements, it is necessary to collaborate with specific teams, such as project managers and IT staff who know the institutional technical settings, such as the repository capacity and internal regulations. In our case studies, the data steward was aware of these rules. In general, data stewards are expected to follow the practices of specific fields, monitor RDM development and good practices, monitor international and institutional laws and policies, and suggest improvements in the institutional RDM workflow. Data stewards can help create DMPs with the necessary detail, anticipate problems with project partners, and monitor the results. Detailed information included in DMPs allows researchers to know where the data are stored and preserved, makes the data more understandable and reusable, and diminishes the risk of data loss, ensuring well-organized management. In other words, good DMPs help researchers to satisfy the FAIR principles for their projects and to plan the management of their data in accordance with the principles. One of the researchers in the case studies confirmed that the collaboration with the data steward was beneficial and that, from then on, no project with their group would start without a DMP in order to avoid difficulties related to the data, such as its organization, management, and ownership.

The application of the collaborative DMP-building method also show that a DMP monitoring process is very important for the establishment of DMP support services at institutions. No matter how detailed a DMP is at the beginning of a project, it may require changes up to the very end of the project. Changes can appear at any time during a project and can lead to undesirable situations and misunderstandings with the project partners. Sometimes, data management rules can change, leading to changes in a DMP, e.g. adding more rules for data backup management, or adding a new responsible entity and new collaborators. One of the researchers observed: *"If we didn't know that the plan should be "live" and constantly monitored, we would never look at it again after its creation, thus missing some important questions"*. A DMP can be created during any stage of a project, even after its completion, with or without approved funding. In all the case studies, the data steward gave the same attention and importance to creating the plan and followed the collaborative method. The one difference was that for projects that were not approved for funding and those that had been completed, monitoring sessions were not scheduled.

DMP monitoring processes are scheduled in advance, after the creation of the first version of a DMP. Thus, some of our case studies require further contact by the data steward with the researchers, since in some cases the DMP is under development yet, and in others it will need to be improved. During the scheduled period, the data steward contacts the researchers and verifies if their project had any changes. We also advise researchers to contact the data steward whenever they need or have any questions related to data management in their projects. The goal is to detect and prevent errors, and register all changes in the DMP, making them available to all people in charge of the project. Taking into account the answers, the researchers considered DMP monitoring very important and would like to be notified for monitoring process by email. The proposed interval (every 6 months) for monitoring session is quite acceptable in the research community; the researchers' responses also point to this interval. The existence of a pre-filled DMP for a specific domain or "a generic template that can be adapted for specific DMP" was also mentioned by researchers in the questionnaire as useful. User support, RDM tools, and good practices were suggested as important for the whole institution. With each new

plan created, both the data steward and the researchers acquire new knowledge and skills and engage more with research data management. In other words, the collaboration and the proposed collaborative method can positively affect several stakeholders at the research organization.

According to the analysis of the maDMP principles, RDM and FAIR best practices and solutions, existing suggestions for the DMP support services and RDM workflow in general, we consider that these processes can be customized and automated separately. This preliminary work also helped us define three profiles where DMP services could have an impact and that would be taken into account in the design, testing and evaluation of the DMP support system for research organization.

The first profile is a regular researcher, who can be a PI or someone in charge of the RDM issues in a project. Using the proposed collaborative method, the data steward creates a detailed DMP for the project. This simplifies the creation of the DMP while also identifying the specific domain issues regarding the DMP. The data steward helps the researcher to keep their DMP "live", monitoring and updating the plan according to a schedule.

The second profile is a researcher who acts as the data manager in a research group. The data steward helps the data manager to create the plan and also provides detailed information about the process. In other words, the data steward provides specific details that improve the RDM skills of the data manager. This also helps them to become a data steward capable of supporting other researchers at the institution. One of the researchers of the case studies confirmed: *"With the knowledge obtained, it was possible to assist a Research Center in reflecting on the need to create internal policies for the management of the data collected within the scope of the projects they develop. And my name was suggested to join a task force to assess issues related to Data Management"*. In this case, the data steward has the role of a consultant who will also be invited to clarify RDM issues and organize workshops or other kinds of training.

The third profile is a collaborator of the institution who in one way or another intersects with research projects and therefore with RDM issues. This could be a project manager, funding officer, DPO, or others stakeholders. The data steward can provide this person with access to the database of the documents created for the project and for monitoring the DMP. This can be useful in various situations, for example for the project manager when necessary to send a report to the funder, or for the DPO to get the first version of the DPIA. In this case, the data steward can be seen as a central element that brings together researchers and the different stakeholders of the institution in the RDM activities.

To improve the proposed collaborative DMP-building method and establish an institutional DMP support system, we continue to collaborate with INESC TEC researchers in a systematic experiment where a cross-section of the projects for a complete year was considered. This is reported in Chapter 7.

5.4 CONTROLLED VOCABULARIES AND FLEXIBLE METADATA IN DMP

During the collaboration with researchers, it was observed that specific aspects of the scientific domain of the project could be identified in the creation of the DMP. Some of these aspects can be seen as elements of the metadata of the DMP that are described during the creation of the plan. In this case, the values used for a specific element of the metadata of the DMP can be used to create controlled vocabularies. They can help to simplify the creation of DMPs and standardize and automate their creation process. They can also help create Domain Data Protocols, which are used as "templates" for plans for specific domains to ensure data quality and comply with RDM requirements and researchers' needs. Domain Data Protocols are "pre-

filled DMPs"[19] that simplify the creation of the DMPs for a specific domain; we described them in Chapter 3, Section 3.4.

To define the controlled vocabularies for DMP, which, in turn, help to create Domain Data Protocols, we apply the collaborative DMP-building method on the projects from the Psychology scientific domain. Previous work with the Faculty of Psychology and Education Sciences of the University of Porto has being continue into two new projects "VigilHate" and "N-DeciMa", and the DMP support to these projects. The person in charge of RDM tasks at the FPCEUP, Mafalda Lopes, that already collaborated with us in her MSc dissertation, defined a metadata model for the Social Science domain that we decided to apply to the projects to show that metadata models and controlled vocabularies could help researchers in the creation of the DMP. This model was proposed during the development of her MSc dissertation "Descrição de dados de investigação: requisitos de investigadores para modelos de metadados na Psicologia e Ciências da Educação" (Research data description: researchers requirements for metadata models in Psychology and Educational Sciences) [160]. It is a result of the analysis of the DDI and the Dublin Core metadata standards and results obtained from researchers during data descriptions experiences on the FPCEUPs' projects. This model was also proposed to attend to the needs of researchers from Psychology and Educational Sciences for research data description. It helped to identify relevant elements in this domain that can be used during the data description and other RDM activities.

Before collaborating with the researchers of the "VigilHate" and "N-DeciMa" projects, was held a meeting with Mafalda Lopes to analyze her proposed metadata model and to identify values of the metadata elements that can be used in the creation of the controlled vocabularies for DMP. We analyzed her metadata model and asked her to detail the values of the controlled vocabularies that she defined in the MSc dissertation, which resulted in an extended metadata model [159]. Table 13 shows different categories of the extended metadata model that can be seen as specific metadata elements of the DMP and examples of the values of these elements that can be used to create controlled vocabularies. We applied this extended model to the "VigilHate" and "N-DeciMa" projects during the creation of their DMPs, along with the collaborative DMP-building method, to show that by using controlled vocabularies, the content of the DMP can be more detailed regarding the specificity of the Psychology domain. We wanted to test if the metadata model and controlled vocabularies help to create Domain Data Protocols, that in turn, help to standardize DMPs in specific domains. Furthermore, we wanted to show that using this approach, it is possible to create metadata models, controlled vocabularies and Domain Data Protocols for other domains while using collaboration with researchers.

5.4.1 VigilHate project

The "VigilHate - Vigilant Citizens Against Hate: How to counter bystander apathy and increase citizens' commitment against online hate speech?" project focuses on the approach to combating online hate speech. This project was funded by "la Caixa"[20], conducted by the FPCEUP at the University of Porto and involved several types of data, including sensitive and personal, that would be collected during five planned studies. The PI of the project had many urgencies with DMP due to requirements by the funder, so they requested our help. They had only two weeks to create and submit the first version of the plan. So, during these two weeks, two meetings were organized, all necessary information was collected, and the first DMP was developed using the collaborative DMP-building method. Moreover, to define controlled vocabularies for this project, we applied the extended metadata

19 https://ddp-bildung.org/
20 https://fundacaolacaixa.pt/

Table 13: Extended metadata model with controlled vocabularies proposed for Social Science domain [159]

Camp	Description	Category	Value of controlled vocabulary
Domain	Identify scientific domain	**Research domain referred to the data**	Classification FOS - Social Sciences - Psychology
Ethics dimensions	Refers to all aspects that indicate which issues highlight ethics in the preparation of a research project for the domain under analysis	**Protection of the participants' identities**	confidentiality process - anonymity (at the beginning) - anonymisation (after the colection of the data) - codification of the participants' identities - data that can identify a person are not requered
		Legal procedures	legal references: - Ethics Committe; - GDPR - DPO
Data Collection and analysis	This category refers to all aspects that define the data collection process where techniques, instruments, types of data collection and data analysis methodology are included	**Techniques and instruments for data collection**	data collection techniques and instruments according to each research domain (in this case psychology): scales questionnaires, interviews (semi-structured and non-directive) , self-report questionnaires , direct observation, ethnographic studies , magnetic resonance , electroencephalogram (EEG), etc. various techniques most of the time combined
		Collection types	types of data recording most commonly used by researchers: Paper Digital
		Methods and techniques of analysis	methodologies and techniques for data analysis Quantitative Qualitative Mixed
Data description	This category refers to the elements referred to by participants for the description of data	**Project**	description of the context of the project giving rise to the data collection: - the author (list of the authors), - the year (numerical VC), - the institution (list of the names of the institutions), - funders (list of the funders) - keywords (e.g thesaurus: APA Thesaurus of Psychological Index Terms or MeSH (Medical Subject Headings) is the NLM controlled vocabulary thesaurus used for indexing articles for PubMed), - definition of the theoretical constructs (identify the scales to which the items correspond, e.g. Creation and validation of a Bystander Effect scale for the Portuguese population, The Domain-specific Risk-attitude Scale, Teste de Rorschach) - number of participants (numerical VC),
		Variables	description and identification of variables: - Condition variables for inclusion/exclusion in the study can be general characterization of the participants (e.g. normal hearing), - Neurophysiological variables (e.g. n3 wave - vestibular evoked neurogenic potential, p1 and p2 waves), - Sociodemographic variables (e.g. nationality, age, profession, level of education, country), - Psychological adjustment variables (e.g. stress, anxiety, sadness).
		Technical characteristics	technical characteristics in the description of the research data: -Format of files (e.g. PDF/A, CSV, TIFF, ASCII, Open Document Format (ODF), XML, Office Open XML, JPEG 2000, PNG, SVG, HTML, XHTML, RSS, CSS, Rich Text Format (*.rtf), MP3, MPEG, JPG, MS Excel (*.xls), SPSS (*.sav, *.por), STATA (*.dta)), - Type of data (e.g. answers to questionnaires, interviews, EEG images, MRI)
Data management and organisation	This category refers to the process of managing and organizing research data	**Data management**	People in charge for data management on the project: - Principal investigator - Supervisor
		Storage local	- work computer, - personal computer, - cloud, - Docolab (collaborative project management software), - Dropbox, - OSF, - Zenodo, - institutional repository, - shared folders on the institutional server.
		Data organization mode	- folders by year, - folders by instrument used, - folders by countries, - folders by type of files, - by project
Data sharing	This category refers to the possibilities of sharing the data (where and with whom the data can be shared)	**Communities for sharing**	- community created on the Zenodo, - OSF, - project website, - funder platform (collaborative for the project)
		Conditions for sharing	- Public - Restricted - Conditional (with access request)
Data storaging	This category refers to the period of time when research data is storaged	**Storage period**	- 10 or 15 years, - 5 years, - embargo until publication, - 2 to 3 after the end of the project, - after the data will be transcripted.
		Licenses	- Creative Commons (http://ufal.github.io/public-license-selector/)
Other additional documents			- informed consent - confidential agreement - authorization from Ethical Committee

Table 14: Questions and answers received from the PI according to the extended metadata model ("VigilHate" project)

Questions	Answers
Identify your domain by the Field of Science and Technology classification (FOS) in the Frascati Manual?	Psychology.
Should your project follow any regulations of the legal bodies like the Ethics Committee, GDPR, and DPO?	Yes, we will send a statement to Ethics Committee, and we will conduct our project according to the GDPR, due to collection of the sensitive and personal data. However, we do not know if we need to contact some DPO.
What type of protection is related to the identity of the participants you will apply?	Sensitive and personal data will be anonymized, codified, pseudonymized, and destroyed as soon as possible.
Which techniques will be used for data collection?	Questionnaires, interviews, and video group sessions.
Will the data be collected in both formats: paper and digital?	No, only digital, and yes, in our domain, we often collect data in paper format.
What method will be used for data analysis?	Mixed (quantitative and qualitative).

model proposed by Mafalda Lopes, interviewing the PI at the first meeting. Table 14 presents questions we asked the PI and the answers we received.

Most of the important questions were related to the list of variables that are very important for this domain. The data steward did not have profound knowledge of this domain and the practices that existed in the domain's projects. Moreover, the data steward did not know about different types of variables, which were important for studies in that domain as well. Thus, we asked the PI about variables and showed the controlled vocabularies list with different metadata elements and their values proposed by Lopes, M.:

– Condition variables for inclusion/exclusion (e.g. general characterisation of the participants such as normal hearing),

– Neurophysiological variables (e.g. n3 wave - vestibular evoked neurogenic potential, p1 and p2 waves),

– Sociodemographic variables (e.g. nationality, age, profession, level of education, country),

- Psychological adjustment variables (e.g. stress, anxiety, sadness).

Analyzing these metadata elements and their values, the PI confirmed that sociodemographic data (gender, age, nationality, education) would be collected in the project. Moreover, the PI mentioned that all projects in their domain usually had different types of variables, highlighting the usefulness of the controlled vocabularies, which could help to remember about those topics during the creation of the DMP.

During the first meeting, we also asked the PI to indicate authors, funders, people in charge, institutions - members of the project, number of participants for each study, format and type of the collected data, local storage, data organization mode, community created for sharing, access right, storage period, and licenses. The information about other additional documents such as informed consent, confidential agreement, and authorization from Ethical Committee was also questioned according to the controlled vocabularies in the metadata model proposed by Mafalda Lopes.

Using the collaborative DMP-building method, the data steward collected information about the context of the project, and organizational questions, evaluated experience level in RDM, motivation of the researchers for RDM tasks, questions, and needs that they faced during DMP creation. As a result of the collaboration, the first DMP was presented for validation and improvements by the PI. After that, the plan was submitted on the "la Caixa" platform[21] and published on Zenodo [140]. Following the steps of the collaborative method, the data steward added this project to the database for monitoring and scheduled the next meeting in 6 months.

One of the specificities of this case was the urgency of the creation and submission of the DMP using the funders' template. So, the collaborative DMP-building

21 https://candidate.lacaixafellowships.org/login

Table 15: Questions and answers received from the PI according to the extended metadata model ("N-Decima" project)

Questions	Answers
Scientific domain by Classification FOS	Psychology (Neuroscience)
Protection of the participants' identity	codification sheet, encrypted and with minimum access; anonymity total (at the beginning)
Legal procedures	Ethics Committee and GDPR
Techniques and instruments for data collection	questionnaires, decisionmaking tasks, self-reports, behavioral decision-making reports, electroencephalogram (EEG)
Methods and techniques of analysis	quantitative
Variables of study 1	Self-reported attitudes towards risk and uncertainty, self-reported psychopathic traits - TriPM, SRP-SH; mean heart rate (HR), Standard Deviations of NN Intervals (SDNN), the Root Mean Square of Successive Differences (RMSSD)
Variables of study 2	cue-RewP and cue-P3 event-related potentials; IUS (uncertainty intolerance) and DRaS (domain-specific risk attitudes) scores, and psychopathic traits measures (TriPM, SRP-SH)
Sociodemographic variables	age, gender, nationality, native language, employment status, level of education, laterality, visual and auditory acuity, mental health and neurological history, medication, substances use (drugs, alcohol); qualityand number of hours slept before the data collection, recent alterations in daily routine

method and the questions, based on the metadata model with controlled vocabularies, allowed us to create DMP quickly and with good detail and quality. After the submission of the DMP on the funder platform, the PI received notification that the plan was approved without any comments or suggestions for improvement.

Another specificity of this case was related to the realization by researchers of the importance and difficulties of DMP creation. This collaboration prompted the PI to include data steward support in the budget of a new proposal under the preparation. Moreover, the timescale for the project deliverable related to the creation, monitoring, and improvement of DMPs extended to the entire project, not just at the beginning or the end.

5.4.2 N-DeciMa project

The second project where we applied the metadata model with controlled vocabularies was "N-DeciMa: Neurophysiological bases of decision-making processes: dissociating risk and uncertainty in the human brain". This project was conducted by the FPCEUP at the University of Porto and financed by BIAL[22] - a pharmaceutical company. Its principal goal was to investigate studies related to experimental dissociating the neuronal correlates of risk and uncertainty processing in decision-making, while controlling for expected value and utility. The project had two studies that would collect different types of data, including sensitive and personal data. Following the collaborative DMP-building method and the metadata model with controlled vocabularies proposed by Lopes, M., we interviewed the PI. After the first meeting, we received all the necessary information for the creation of the first DMP draft. We also analyzed the project proposal document and statement submitted to the Ethics Committee.

The use of the metadata model with controlled vocabularies proposed by Lopes, M. gave us specific information related to the project and allowed us to identify some values for metadata elements such as variables (Table 15).

In this study, along with variables, it was also important to identify scales used during experiences. *The Domain-specific Risk-attitude Scale* [23] and *the Intolerance of Uncertainty Scale* [24] are two scales adapted for European Portugueses that help researchers to obtain data for their projects. The PI confirmed that most projects in their domain used variables and scales. Thus, the list of controlled vocabularies with the different values of the scales elements in the metadata model that we presented to the PI helped us not to forget to include this topic on DMP and explain it in more detail.

22 https://www.bial.com/com/
23 http://repositorio.uportu.pt/xmlui/handle/11328/4521
24 http://repositorio.uportu.pt/xmlui/handle/11328/4520

We also asked the PI about other types of values that can be used to create controlled vocabularies to describe their project. In this context, we obtained two important lists that are registered on the Open Science Framework page of the project laboratory [25] and are constantly used by researchers of this project:

1. Sample of the values to be used in the controlled vocabularies, referring to questionnaires and scales[26]:
   - TriPM - Triarchic Psychopathy Measure;
   - STICSA-T - State-Trait Cognitive and Somatic Anxiety Scales;
   - IPAS - Impulsive Premeditated Aggression Scale;
   - BPAQ - Buss Perry Aggression Questionnaire;
   - BSI - Brief Symptom Inventory;
   - DERS - Difficulties in Emotion Regulation Scale;
   - HADS - Hospital Anxiety and Depression Scale;
   - BIS-BAS - Behavioral Inhibition/Behavioral Activation Scales;
   - YPI - Youth Psychopathic Traits Inventory;
   - ACE - The Adverse Childhood Experiences.
2. Sample of the values to be used in the controlled vocabularies, referring to protocols[27]:
   - DotProbe - Protocol: DotProbe for EEG/ERP;
   - Logos - Protocol: Logos for EEG/ERP;
   - Butterflies - Protocol: Butterflies behavioral task for EEG/ERP;
   - Wheel of Fortune - Protocol: Wheel of fortune task for EEG/ERP;
   - Affective Faces - Protocol: Affective faces task for EEG/ERP;
   - Bandit Neutral Harm - Protocol: Bandit neutral harm task for EEG/ERP;
   - Inverted Faces - Protocol: Inverted faces task for EEG/ERP;
   - Auditory Oddball - Protocol: Auditory oddball task for EEG/ERP;
   - Oddball Dual Task - Protocol: Oddball dual task for EEG/ERP;
   - Infants Ratings - Protocol: Infants rating task for EEG/ERP.

Although their plan was not published yet and was in the review stage by the PI, they showed interest in publishing it. Moreover, by creating the plan with the support of the data steward, they were improving the way of managing sensitive and personal data, diminishing the risks that may exist. Following the collaborative DMP-building method, this project was added to the monitoring.

5.4.3 Proposal of the controlled vocabularies and Domain Data Protocols for Psychology scientific domain

These case studies show that besides the collaborative DMP-building method, the use of the metadata model and controlled vocabularies also help in DMP creation. They make the process easier for both the data steward and the researcher. They can help the data steward to become familiar to specific domain, also serving as a reminder for important topics that need to be described in the DMP. The controlled vocabularies could help the data steward to make the questions for researchers more specific, suggesting different variants of the terms. This, in turn, helps researchers focus on the project's important details and reduces the time taken by the data steward to gather information to create good-quality DMP.

Controlled vocabularies can standardize the plan for a specific scientific domain, and can be expanded and adapted for similar domains. Such terms as recorded sessions, video group sessions, simulation games, and decision-making tasks, which we have heard during our collaboration, could be added as new terms for controlled vocabulary. For the controlled vocabularies, which describe licenses, more values of the licenses could be added; for additional documents, more values of the documents, such as *Data Protection Impact Assessment*, could be added.

25 https://osf.io/8wd8m/
26 https://osf.io/8wd8m/wiki/Questionnaires/
27 https://osf.io/8wd8m/wiki/Protocols/

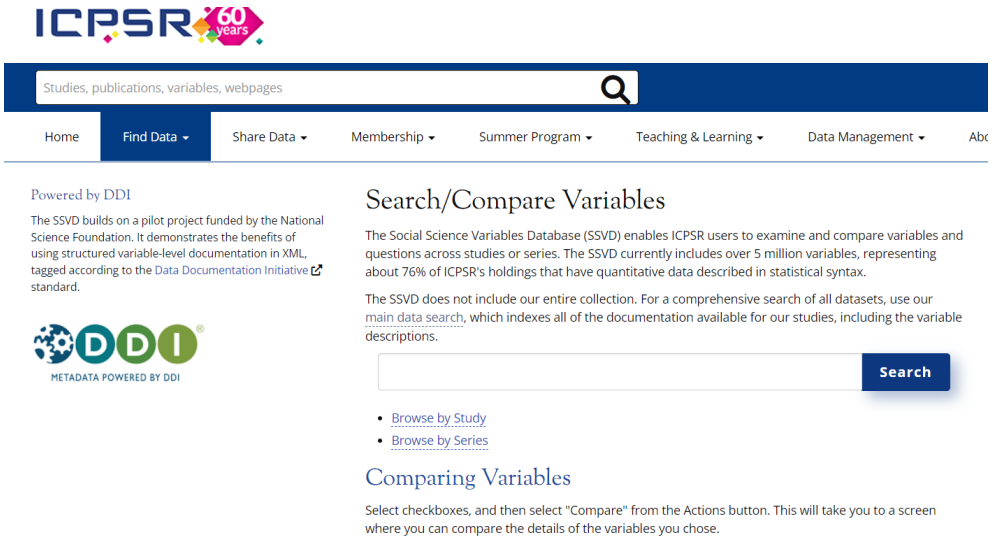

Figure 53: Interface of the ICPSR repository for variables

The importance of the variables for this domain is also highlighted by the ICPSR data repository, which is one of the main repositories for depositing data results for these domain projects. The ICPSR has a separate place for variables (Figure 53), where researchers can find or compare them.

When researchers create a DMP, they often do not know what information they need to indicate, so metadata models and controlled vocabularies could provide guidance for DMP creation. Domain Data Protocols, described in Chapter 3, Section 3.4 were also seen as a guide for specific domains. In our cases, using "A Framework for Domain Data Protocols" [73], the analysis and application of the metadata model with controlled vocabularies allowed us to make the suggestions for Domain Data Protocols for the Social Science domain. As protocols generally include different types of questions, in Table 16 we present our suggestions for this domain in the form of the questions/answers.

Finally, these projects also highlight the important role of the "third profile", defined during the application of the collaborative DMP-building method, as *".. is a collaborator of the institution who in one way or another intersects with research projects and therefore with RDM issues..."*. The person in charge of RDM tasks at the FPCEUP helped us in the DMP creation process, improving her specific knowledge related to the RDM issues. The collaborative method can, therefore, be used as a base or example for developing their own DMP support system.

5.5 SUMMARY

This chapter describes the collaborative DMP-building method that was developed to support researchers in DMP creation and monitoring processes. This collaborative method was applied in different case studies and results of this application were analyzed. The results show that the data steward can help researchers in DMP creation at different stages of the project. The DMP creation and the subsequent monitoring process requires a lot of effort and time to result in a detailed and published DMP. From the creation of the plan to its publication, the data steward organizes different meetings, clarifies all RDM issues, analyzes information related to the project, and collaborates with other stakeholders when needed. After collecting and analyzing the necessary information, the data steward proposes the first draft of the DMP and, after its correction and improvement by the researchers, publishes

Table 16: Suggestions for Domain Data Protocols for the Social Science domain

Questions	Examples of the answers (controlled vocabularies)
1. Identify your scientific domain by FOS	a. Social Science / Psychology; b. Social Science / Sociology; ...
2. Indicate techniques/instruments for data collection	a. Questionnaires; b. Interviews; c. Self-reports; d. Direct observations; ...
3. Choose questionnaires for your researcher	a. TriPM - Triarchic Psychopathy Measure; b. STICSA-T - State-Trait Cognitive and Somatic Anxiety Scales; c. IPAS - Impulsive Premeditated Aggression Scale; d. BPAQ - Buss Perry Aggression Questionnaire; e. BSI - Brief Symptom Inventory; ...
4. Indicate methods for data analysis	a. Quantitative; b. Qualitative; c. Mixed.
5. Describe variables that you will use in your project	a. Sociodemographic variables (e.g. nationality, age, profession, level of education, country); b. Neurophysiological variables (e.g. n3 wave - vestibular evoked neurogenic potential, p1 and p2 waves); ...
6. Choose protocols that will be used during the project	a. DotProbe - Protocol: DotProbe for EEG/ERP; b. Logos - Protocol: Logos for EEG/ERP; c. Butterflies - Protocol: Butterflies behavioural task for EEG/ERP; d. Wheel of Fortune - Protocol: Wheel of fortune task for EEG/ERP; ...
7. Please indicate scales that will be used on the project	a. The Domain-specific Risk-attitude Scale; b. The Intolerance of Uncertainty Scale; ...
8. Do you need support from the DPO?	a. Yes; b. No.
9. Please indicate additional documents that you need to create for your project	a. Statement / authorization from Ethics Committee; b. Data Protection Impact Assessment; c. Informed consent; d. Confidential agreement; ...
10. Software used for data analysis	a. Matlab; b. Qualtrics; c. BIOPAC; d. IBM SPSS; ...
11. Please indicate the way of the protection of the identity of participants	a. Anonymity (at the beginning); b. Anonymization (after the data collection); c. Codifying of the identity; d. Pseudoanonymization; ...
12. Please indicate responsibilities on the project	a. For DMP; b. For data collection; c. For data preservation; ...

it and also controls the timing, sometimes insisting with researchers regarding this task.

Preliminary results also show that the collaborative method is suitable for projects from different scientific domains and can be used in different institutions. Using this method, the data steward facilitates the creation DMP process and its monitoring, making plans with more quality and detail. Researchers, in turn, diminish their effort and time for DMP, clarify all their issues regarding the RDM in general, do their projects in compliance with RDM requirements and have support at any stage of the project.

In this chapter, we also apply the collaborative DMP-building method to projects from Psychology scientific domain, together with metadata model and controlled vocabularies proposed for Social Science domain. We show that metadata models and controlled vocabularies make DMP creation easier and can be seen as a useful tool during the interviewing researcher. They can be used as guides to remind data steward to discuss the specific points of the domain or to suggest different variants of the terms, making DMP with more detail and quality. The projects for the Psychology domain described in this chapter show the way to define controlled vocabularies for other scientific domains according to their specificity. Controlled vocabularies can be used in different elements of the DMP support system during its implementation. We also suggest the Domain Data Protocol for the Social Science domain as an example of the standardization of the plans for projects of a specific domain.

6 INTEGRATION OF A DMP WORKFLOW AT THE INSTITUTION

To establish an institutional DMP support system, we should analyze different aspects of the institution. Definitely, we need to continue to collaborate with researchers from different domains in their projects to improve and systematize the proposed collaborative DMP-building method. But at the same time, we also need to analyze other specific aspects, such as the project management process existing in the institution, tools, and platforms, to understand how the developing DMP support system could be implemented. Moreover, we also need to be aware of the evolution regarding the DMP, best practices and recommendations.

In this context, this chapter focuses on the analysis of the project management process and internal regulations at INESC TEC, identifying the way to integrate the DMP workflow at the institution. Moreover, in this chapter, we look at the possibility of making the developing DMP support system in line with the maDMP standard proposed by the RDA DMP Common Standards Working Group[1].

The proposal for the interface of the DMP support system is also suggested in this chapter. We suggest its implementation on the INESC TEC Research Information System (IRIS) platform [2], used at INESC TEC for project management, demonstrating how DMP workflow can be presented and how researchers can request DMP support from the data steward from the beginning of their projects. We also show how this implementation allows automating DMP-related mechanisms, simplifying the work of the data steward to keep the researchers' DMPs "live", verify the quality of the DMPs, edit or add more information, contact the researchers when required, validate plans with the DPO, and collaborate with other stakeholders when it is necessary.

Some of the part of this chapter is based on the published report: Karimova, Y., et al. "Research Data Management Workflows and maDMPs", 2020 [141].

6.1 ANALYSIS OF THE PROJECT SUBMISSION PROCESS AT THE INSTITUTION

The internal procedures, workflows, processes and tools used to support researchers in project management play an important role in the RDM workflow. Our collaboration with the Business Informatics Service, Data Manager Department, Funding Officer and DPO allows us to analyze the current situation of the project management process at the institution to define how the developing DMP support system could be implemented and what we need to propose for this implementation.

The meetings held and analysis of the project management process shows that the proposals of the projects are created on the INESC TEC Intranet (internal management platform)[3] (Figure 54).

In the normal project flow (Figure 55) a proposal is created by the researchers who want to submit the proposal. Then this proposal automatically goes to the person in charge of the center or service to which the researcher is allocated for approval. In the next step, the proposal is controlled by the Serviço de Apoio à Angariação de Financiamentos - Fundraising Support Service (SAAF) where the budget for the

1 https://www.rd-alliance.org/groups/dmp-common-standards-wg
2 https://iris.inesctec.pt/
3 https://intranet.inesctec.pt/

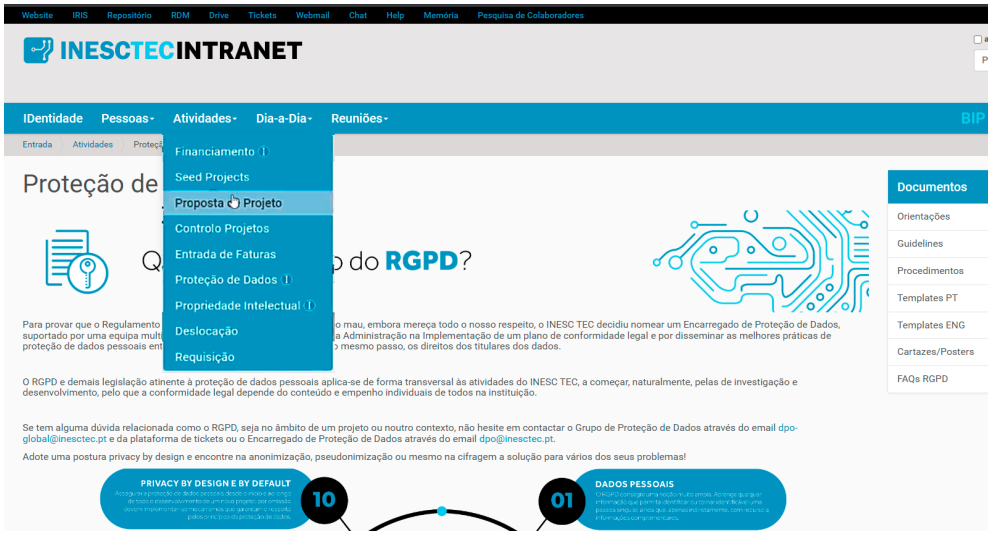

Figure 54: Interface of the INESC TEC Intranet

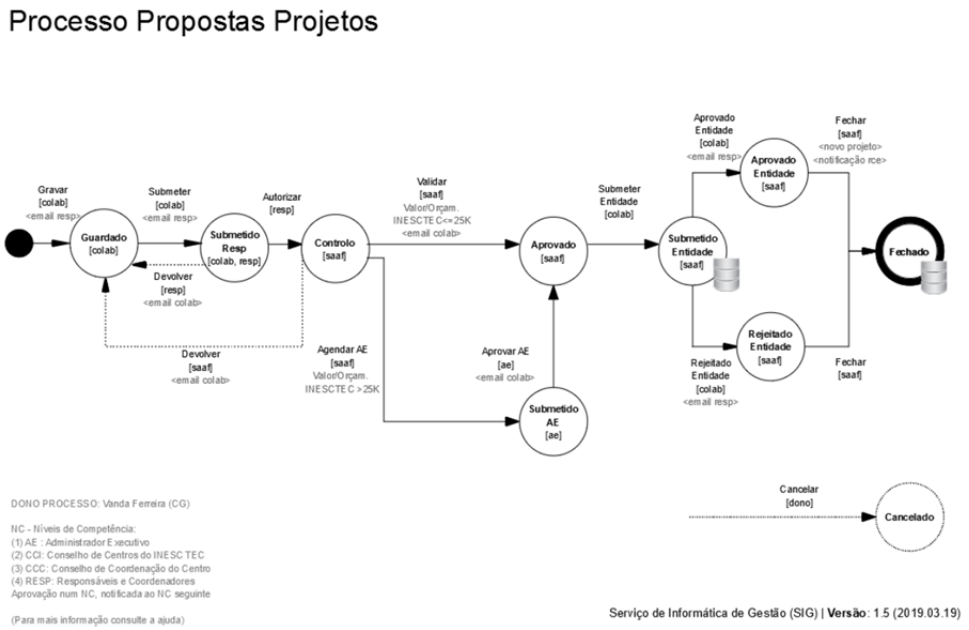

Figure 55: Project proposals Workflow

proposal is verified. If it is higher than 25k, it must proceed through an Executive Director; if it is lower, it is approved directly.

After this, the proposal is sent to the client or funder for being *accepted* or *rejected* for funding, continuing to be on the INESC TEC Intranet platform in the mode *awaiting the response*. When the proposal is approved or rejected, the PI informs the SAAF, in the case of international projects, or the Controlo de Gestão - Management control (CG), in case of national contracts. They check if everything is complete and decide to *close the process with continuation* (i.e. proposal accepted) or *close the process with cancelling of the proposal* (i.e. proposal rejected) (Figure 56). When an application is approved for continuation, the project record automatically enters the database on the IRIS platform for follow-up. Projects entering on IRIS are always created on the basis of proposals on the INESC TEC Intranet. During the transfer process, all information from the INESC TEC Intranet is transmitted to IRIS for further revision and supplementation with various documents, agreements and other important information.

Figure 56: Example of the canceled proposals

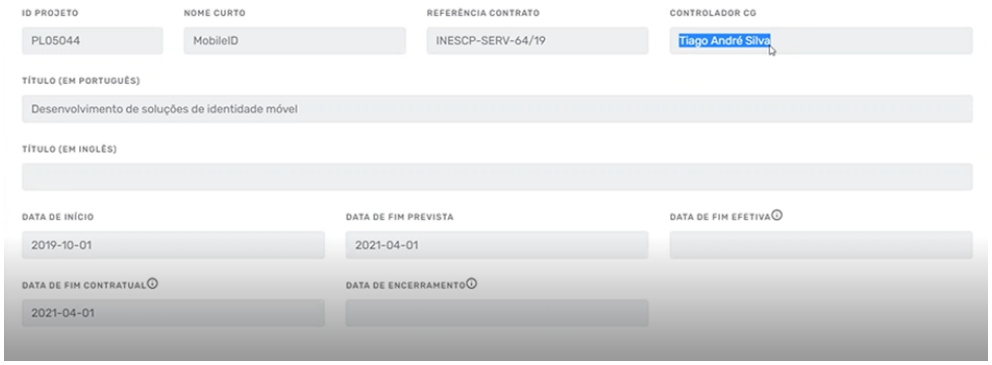

Figure 57: Information about the project on IRIS

After the project is on IRIS, the CG controller (see Figure 57) and the project officer can edit and improve the project information (for example, add associated documents, upload them by the Management Support (GA), edit data, metadata, and add more descriptions related to the project). Some metadata (e.g. project ID) is generated automatically by the Serviço de Informática de Gestão - Management Information Systems Service (SIG); the other comes directly from the proposal (e.g. short name of the project).

The proposal continues to be on the INESC TEC Intranet only as a reference with the project on IRIS and is visible to the person in charge of the center and the author of the proposal. All documents, information and management of the process are coordinated on IRIS.

It is possible to request the support of a DPO during the creation of a proposal on the INESC TEC Intranet. The DPO at INESC TEC assists researchers with GDPR issues related to personal and sensitive data. During the filling of information about the project (Figure 58), researchers have to respond to questions on the DPO tab (Figure 59). If the researcher chooses the option that the project will deal with sensitive or personal data after the proposal has been approved, the DPO team receives an email with a support request automatically. However, it can be configured differently, for example, by receiving an email automatically, even during the proposal.

On the DPO side, a list of all the projects that have requested support can be seen for their processing.

In general, analysis of the proposal workflow and project management at INESC TEC shows that researchers do not know where they can obtain help in the DMP creation. Although INESC TEC provides data stewardship support as an experimental service, the institution does not have any official service, process or tool implemented in its project management process yet, as implemented for the DPO support, for example. So, sometimes, researchers create DMP without any support facing many difficulties.

Based on this analysis, we can propose our version of a strategy for the integration of the DMP workflow at the institution.

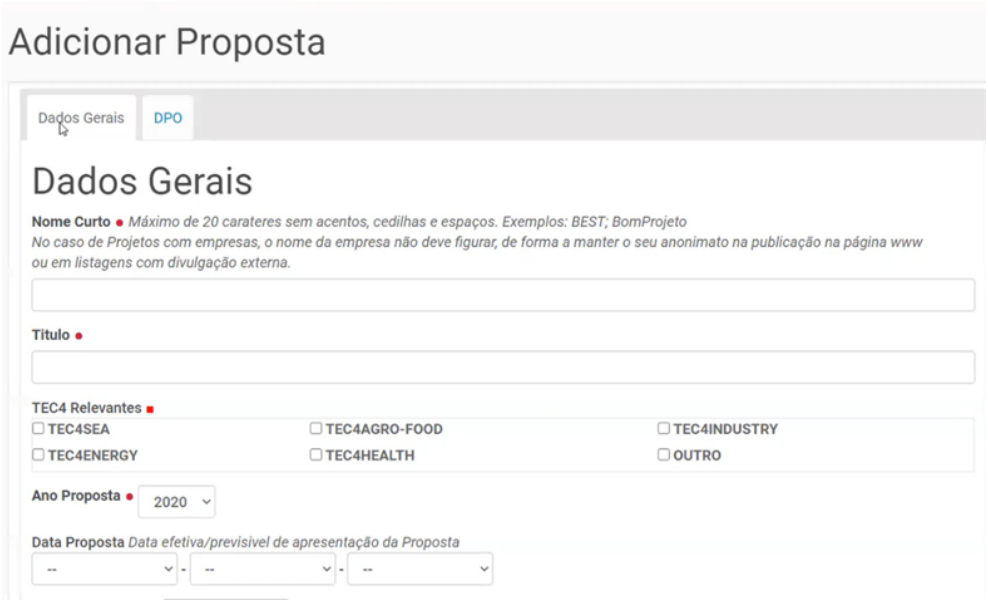

Figure 58: Data to be filled in on project proposals on the INESC TEC Intranet

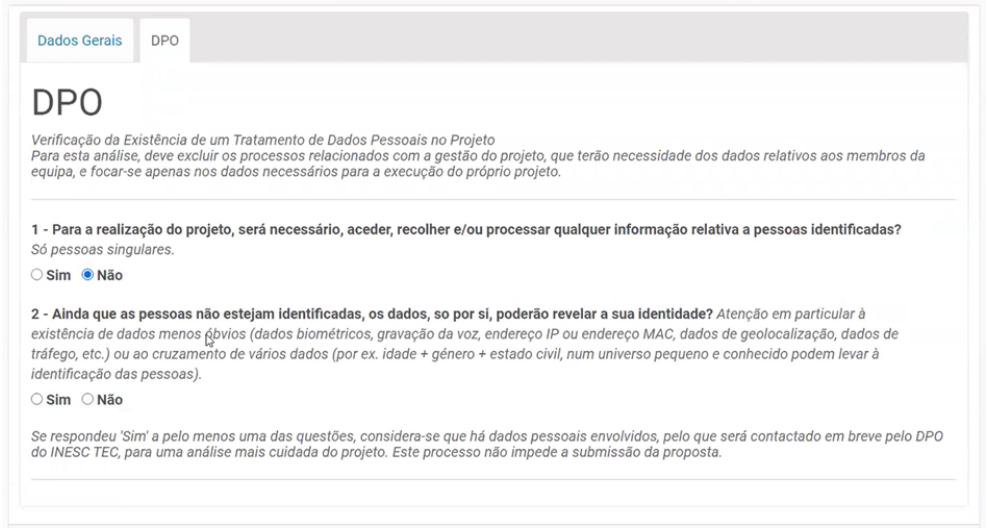

Figure 59: Information related to personal and sensitive data

6.2 ORGANIZATIONAL DMP WORKFLOW AND THE MACHINE-ACTIONABLE DMP

Since the main agencies started to require DMP to be included in grant applications, institutions started to develop and implement or improve existing RDM workflows to support their researchers in various tasks, including DMP creation. The DMP-building method described in Chapter 5 was the first attempt to establish a DMP workflow for research projects and to ingrain RDM into project activities. However, to improve on the DMP support integration, we also need to analyze its possibility of being in line with the machine-actionable DMP standard. The implementation and development of systems based on the maDMP concept, described in Chapter 3, Section 3.2 are one of the recommendations by the RDA DMP Common Standard WG for the research community. The aim of this recommendation is to standardize DMPs and tools to their creation, helping in exchanging of plans between platforms and systems, and facilitate and automate the DMP creation process.

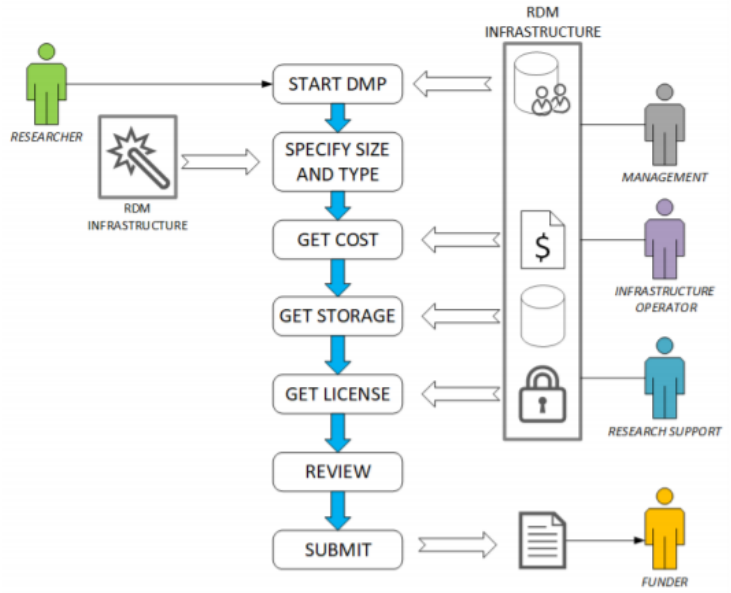

Figure 60: Scheme of DMP Workflow proposed by RDA DMP Common Standard WG [172]

In order to ensure the DMP workflow is in line with the maDMP standard, a number of tasks should be carried out to implement this concept. In more detail, it is necessary to analyze the current DMP workflow state to identify the points that should be improved according to the maDMP, to determine the improvements, implement them, and test them.

The analysis of the INESC TEC DMP workflow state started during the RDA Hackathon on maDMPs[4] organized by the RDA DMP Common Standard WG in 2020. During the hackathon, the work of our group "InsTmaDMP"[5] was focused on the analysis of existing RDM and DMP Workflows from different institutions, namely the INESC TEC, the FPCEUP, the IPO-Porto, and the George Washington University (GWU). We analyzed the DMP creation process at these institutions and compared our methodologies with the maDMP concept to identify the DMP elements that require changes (e.g. addition, elimination, and edition) to make our DMPs machine-actionable. We identified several requirements to help us develop our institutional DMP Workflow to conform to the maDMP concept [141]. Although we have carried out this work for the above-indicated institutions, we described it in detail only for INESC TEC [6].

First, we analyzed the scheme of the typical process for creating a DMP as proposed by RDA DMP Common Standard WG (see Figure 60), which presents the workflow for DMP creation and stakeholders involved at different steps [172], and created a similar scheme for INESC TEC in the current state (see Figure 61).

This first step helped us to understand how the DMP workflow is organized at INESC TEC and which stakeholders are involved. Since our institution does not have support for researchers to create DMP as a service or system official (only experimental service), in most cases, the researchers create a plan without any support and collect all necessary information on their own. However, they can obtain support from the DPO and their team on issues related to the GDPR (e.g., ethics issues, issues related to sensitive, personal data). According to the proposed collaborative DMP-building method, a data steward can help researchers in DMP creation during all phases. The data steward can be a binding link between researchers and

4 https://github.com/RDA-DMP-Common/hackathon-2020
5 https://github.com/RDA-DMP-Common/hackathon-2020/blob/master/results.md#instmadmp
6 https://tinyurl.com/ycwydo4a

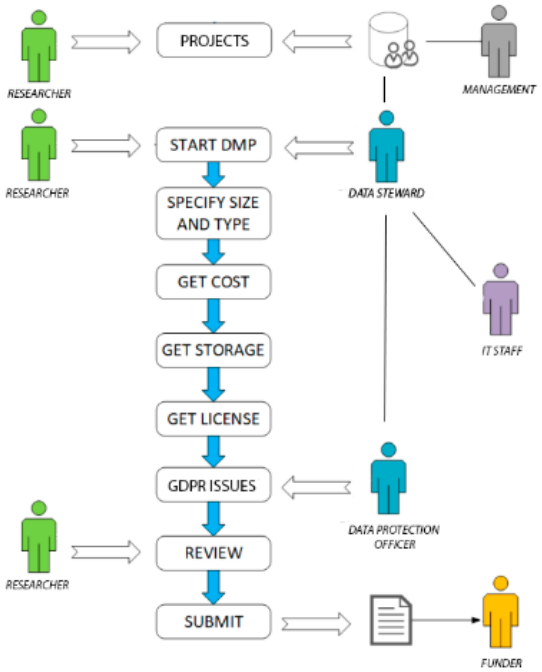

Figure 61: Scheme of the current DMP Workflow at INESC TEC

the IT, DPO departments or other stakeholders existing in the institutional RDM infrastructure.

As a result of the analysis of both schemes presented as a DMP Workflow (Figure 60, 61), we identified points that should be improved:

1. Develop and implement the DMP support system as a service at INESC TEC, making the data steward support a part of the RDM infrastructure and thereby making the process of the DMPs creation and their monitoring more organized, clear, and official;
2. Include the data steward as a stakeholder in INESC TEC's RDM infrastructure, providing a formal point of contact for researchers to support DMP and other RDM issues at the beginning of the project;
3. Bring all stakeholders together to create DMPs with good quality and detailed descriptions that meet the RDM and funders' requirements;
4. Develop and implement mechanisms that allow rendering the DMP support system automated and standardized, according to the DMP and maDMP principles;
5. Create guidelines and documentation for the operation of the DMP support system and services at the institution.

To understand how we can change our DMP Workflow to have our DMP machine-actionable, we also analyzed the maDMP diagram[7] (Figure 62) and maDMP application profile structure[8] (Figure 63) that represents the fields and elements used on the maDMP standard.

We created our scheme equivalent to the maDMP diagram for the following comparison and mapping of the elements of the standard with the elements of our current structure for DMP creation, according to the proposed collaborative DMP-building method applied on the INESC TEC (see Figure 64).

As a result of comparing the INESC TEC scheme (Figure 64) and maDMP structure scheme (Figure 63), we have identified elements that could be improved to

7 https://tinyurl.com/mu9utv6p
8 https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard

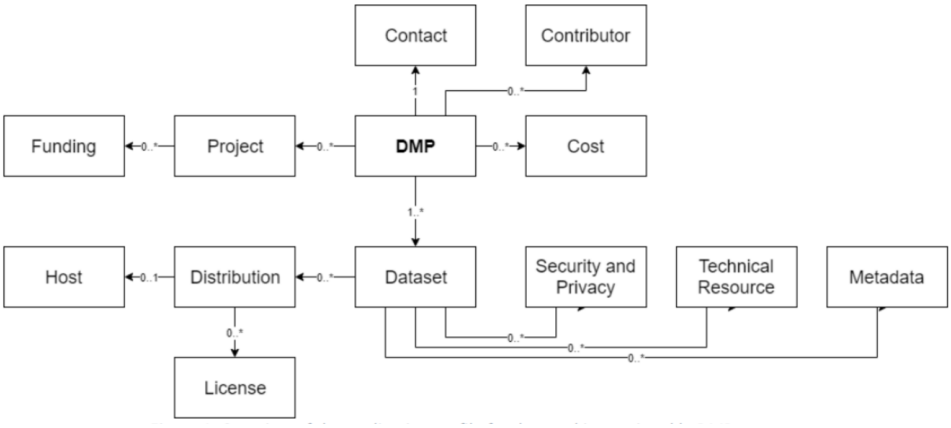

Figure 62: maDMP diagram (Scheme of DMP content) [69]

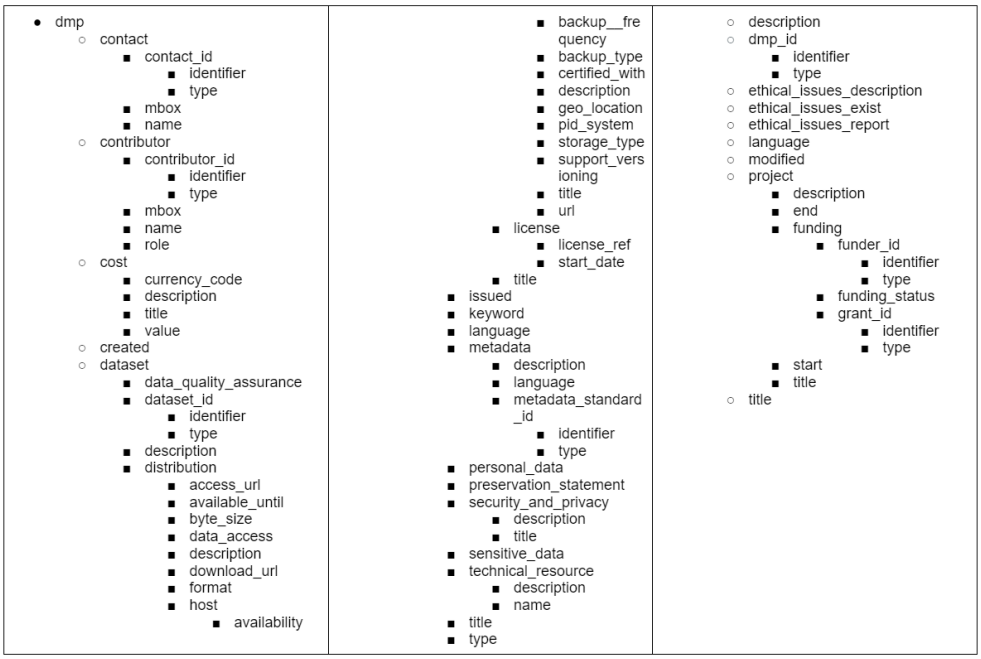

Figure 63: maDMP structure [69]

turn our DMP in line with the maDPM standard, mapped them and proposed the improved DMP structure at INESC TEC (Figure 65).

The proposed DMP structure is based on the maDMP structure and extended with important elements from the INESC TEC DMP structure. These elements were identified during the collaboration with institutional researchers by the data steward while supporting DMP creation. More specifically, we added the elements such as dataset-distribution-available since, dataset-distribution-responsible, dataset-restriction, dataset-documentation, dataset-personal data, dataset-private data, dataset-preservation statement, dataset-tools (see Table 17). These elements are important for INESC TEC researchers and should be used by them during the DMP creation. For example, element *legislation compliance* can be used when researchers need to show what additional documents and/or actions are required to manage the sensitive, personal or private data (e.g. DPO involvement or authorization from the Ethics Committee).

The new DMP structure was tested by two researchers who had already created their DMP for their projects following our old DMP structure. One plan was

Figure 64: DMP structure at INESC TEC

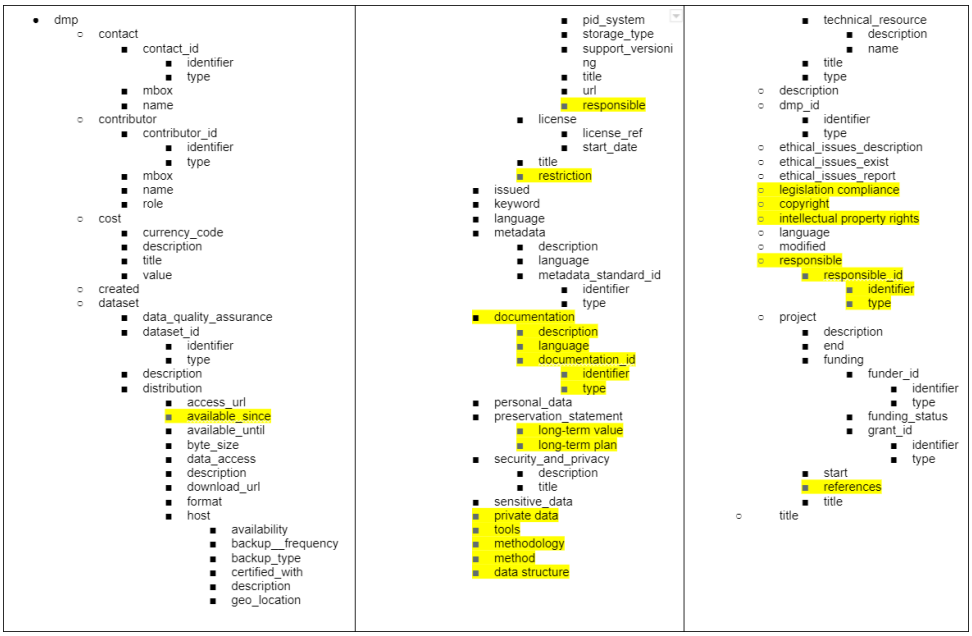

Figure 65: Improved DMP structure at INESC TEC according to the maDMP standard

created for "Gamma Radiation Monitoring" project[9],[10], another one for "FARSYD project"[11] [158]. Creation of the new versions of their existing DMPs based on the new DMP structure, allows us to compare both versions of the plans and analyze the proposed elements for improving DMP process creation. Moreover, it allows us to collect feedback from the researchers and their opinion about the new DMP structure. Thus, we received the following opinions: *"The process of recreating the DMP using the new structure was quite difficult, as the meanings of some fields were not well understood, even with some additional description. In the first view, when doing DMP using a new structure, it leaves some information, which I had in my DMP, outside. And*

9 https://www.arm.gov/research/campaigns/ena2015grm
10 https://tinyurl.com/vusz7yca
11 https://tinyurl.com/2f4hw6kw

Table 17: Elements of the improved DMP structure according to the maDMP standard

Element	Description
dataset - distribution - available since	It is necessary for cases when a dataset has some embargo period. This element could demonstrate when the dataset starts to be available to others.
dataset - distribution - responsible	The element "responsible" serves to identify the person in charge of the deposit and description of the datasets. It could help other researchers establish contact with the correct person if it is necessary.
dataset - restriction	In case when a dataset has some restrictions related to its openness, this element could help to describe what type of restrictions exist and how they could be resolved.
dataset - documentation (description, language, documentation_id (identifier; type))	These elements are considered important for description in DMP because in almost all projects exist different types of documentation (e.g., data documentation that accompanied the data).
dataset - personal data	This type of data can also be created/produced during the project and can require different type of licensing and managing rules.
dataset - preservation statement (long-term value; long-term plan)	These elements can help to understand how long the dataset will be preserved and where it will be stored.
dataset - private data	Along with public, sensitive or personal data, private data can also be created/collected during the project. These data may contain technology or business secrets and may require specific management rules, for example, the creation of the specific agreement between interested partners. Thus, it is important to mention and describe them in DMP.
dataset - tools	This element can help to indicate tools used during the project. Some different tools vary for the scientific domains and can be presented as controlled vocabularies.
datasets - methodology and method	Here the methodology and method for data collection are further detailed and described.
dataset - data structure	The organization of datasets also requires much attention at the beginning. For example, datasets can be organized by tools, by year or by localization, and during the project, it will be more difficult to change than to sort it out at the beginning. It is truly important to plan this information at the very start of the project, namely during DMP creation.
legislation compliance	This element will show what additional documents and/or actions are required to manage the sensitive, personal or private data. For example, such data often requires DPO involvement or authorization from the Ethics Committee.
copyright and intellectual property rights.	These elements contain information that could help to define copyright matters and intellectual property rights for project members.
responsible (responsible_id (identifier; type))	Responsibilities should be clearly defined during the project to avoid confusion and time-consuming searching for those who are responsible for a particular task. Different persons or entities may be responsible for different RDM tasks, for example, one - for the DMP creation, others - for data preservation.
project - references	Since the DMP is often perceived by researchers as the output of a project and is advised for publication with the DOI and citation, it is viable to add bibliographic references to it.

also, I have doubts that this type of DMP is informative in cases of domains like biodiversity since it seems to me to be much more aligned with less heterogeneous data domains"; *"I didn't feel that using the new structure will lack a lot of information if compared to my first DMP. But the way the items are organized in the structure is difficult to understand and I needed the detailed description of the meaning of the fields. In the first view, I think that some new fields that have improved structure make sense to exist, as well as information on the certification of the repository, as I think this information is relevant and should be in the DMP from the beginning. The previous version of the DMP seemed to me to be a task for the PI, this DMP seems to be done by someone on the side of the data deposit and not the research itself"*.

Based on the results and feedback from researchers, we can state that both DMP structures help researchers in DMP creation. However, the one already used in INESC TEC is more understandable to researchers, using more human language; the other is aimed for machines, using more technical elements. This results from turning our DMP into a machine-readable format. The collaborative DMP-building method and structure used in INESC TEC are tailored to the needs of researchers and have been developed to meet the difficulties they face in creating a DMP. When assisting researchers, the data steward always explains in more detail the meaning of each aspect to be described in the plan; the data steward is, therefore, very important in the implementation and configuration of the maDMP in the RDM workflow.

The results also show that although the new DMP structure seemed complicated to the researchers, it included the most important points to be considered when creating a DMP. Researchers indicated that the new DMP structure has several important elements that need to be planned since the beginning of the project (e.g. certification of the repository).

6.3 PROPOSAL OF THE DMP WORKFLOW AT INESC TEC

The analysis of the institutional project management process, internal rules and stakeholders, and work developed at the Hackathon is the first attempt to integrate DMP support at the institution and make the DMP workflow according to the maDMP standard.

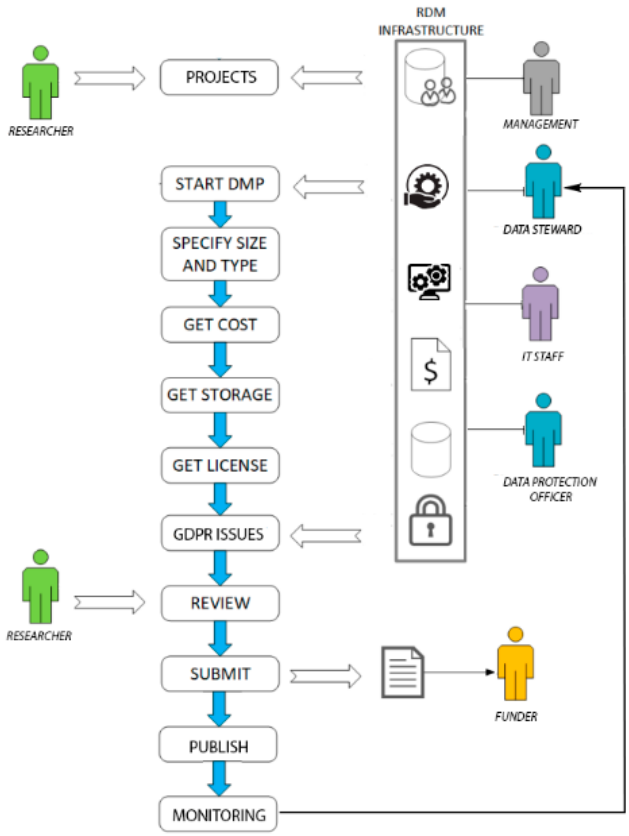

Figure 66: Proposal of improvements for DMP Workflow at INESC TEC

As one of the results of this work, we propose an improved INESC TEC DMP workflow in accordance with the maDMP standard (Figure 66). It covers all important stages of DMP creation, its monitoring, and the stakeholders involved in these processes. The data steward, in this case, plays an important role, helping researchers in all stages of the DMP creation and communicating with other stakeholders and institutional departments when it is necessary, always monitoring the submitted and published DMP.

To make the proposed DMP structure (Figure 65) more understandable for researchers, we transformed our structure into a list of topics that should be filled during DMP creation (see Table 18). This list could be used during meetings with the researchers to support them in the DMP creation. The descriptions of the elements taken from the maDMP structure, which were added to our structure without any change, are taken from the RDA DMP Common Standard website [12].

Summarized and analyzed all the collected information related to the DMP support system and workflow at an institution (e.g., controlled vocabularies, maDMP standard, new DMP structure, improved DMP workflow) we designed a new scheme of the project proposal process (Figure 67). This scheme presents how the project submission process can be implemented into the existing RDM workflow at INESC TEC.

One of the most viable changes we suggest is to incorporate DMP and DPO support activities into the existing project process.

The first DMP request can be realized by the researcher during the project submission phase of the grant application. The practice shows that most funding agencies require the first draft of the DMP to be submitted with the project proposal, so the researcher and the institution should have a DMP support system to help with this task.

12 https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard#structure

Table 18: List of the topics that should be filled during DMP creation

Nº	Topic
1	Title of the project
2	Information about the project (description, start and end dates, funder and funding status, grant number)
3	Contact person (name, email)
4	Author and contributor of DMP (name, email, type of contributor (e.g. data steward), ORCID)
5	Information about datasets: a. type of dataset; b. information about sensitive data; c. information about personal data; d. information about private data; e. description of each dataset; f. how and where dataset will be deposited (e.g., access rules, DOI or URL, dates of the availability (since and util), size, format, repository, storage and backup rules); g. information about licenses and restrictions if exist; h. metadata standard used for dataset description; i. documentation; j. information related to data preservation; k. security and privacy issues; l. tools, methods, technical resources and methodology used during the project; m. data quality assurance.
6	Ethical issues and legislation compliance (e.g., GDPR issues, DPO collaboration)
7	Costs for RDM required during and after the project (currency code, value, description(e.g. costs for storage and backups))
8	Responsibility entity and person in charge of RDM (e.g., responsible for DMP, for data collection, for preservation, for backups)
9	Intellectual Property Rights and copyrights
10	Language of the DMP expressed using ISO 639-3
11	Dates of DMP creation and modifications
12	References

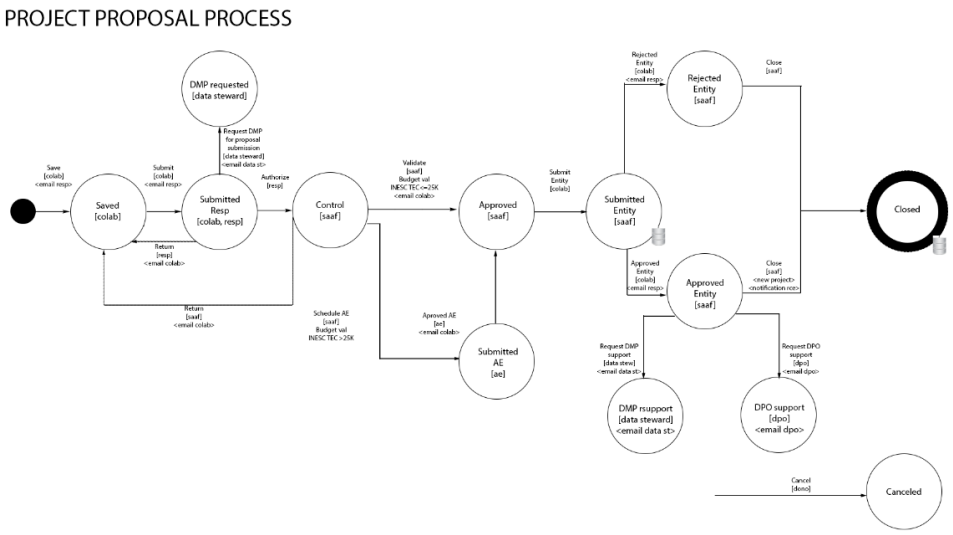

Figure 67: Improved project proposal process at INESC TEC

The second DMP and DPO support requests can be realized when the project is approved by the funder. In this case, the data steward and the DPO team receive notification by email and then contact the researchers to assist them in creating a more detailed version of the DMP; as well as to analyze the risks of sensitive, personal and private data and, if necessary, to assist in the creation of additional documents, such as the DPIA.

In addition, the inclusion of the DMP support system in the RDM workflow at INESC TEC could allow researchers to consult the data steward at any other phase of the project from its planning to its completion, as it could become a service of the institution. This, in essence, could also help increase the reputation of the institution in its assessments for funding, demonstrating that the institution is interested in being compliant with all RDM requirements not only for research projects but also for the institutional infrastructure.

The analysis of the project submission process and the request process for the DPO support presented above (Figure 68) help us to propose, in a similar way, a DMP tab (Figure 69), where researchers can request DMP support for their project during the creation of a proposal. In other words, the creation of the project pro-

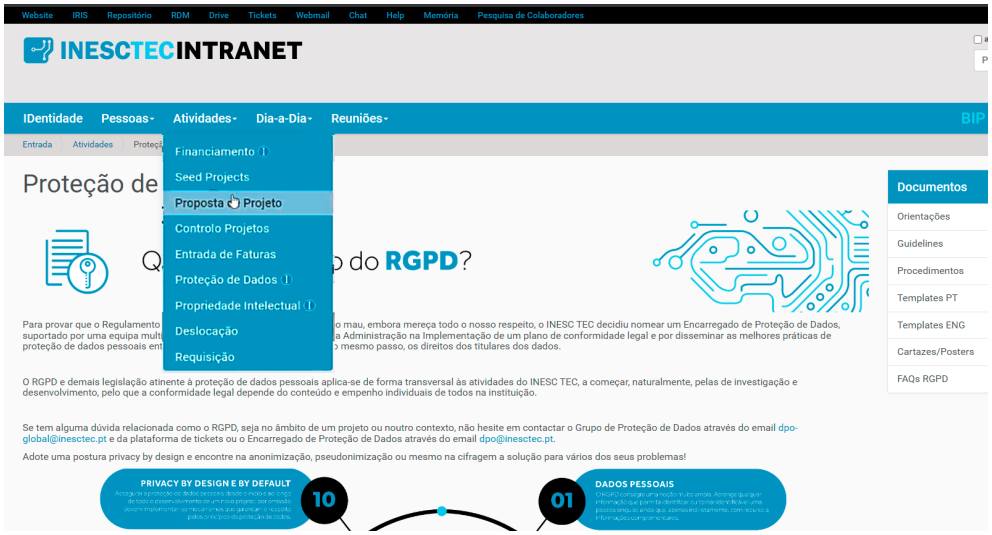

Figure 68: Creation of the project proposal on the INESC TEC Intranet

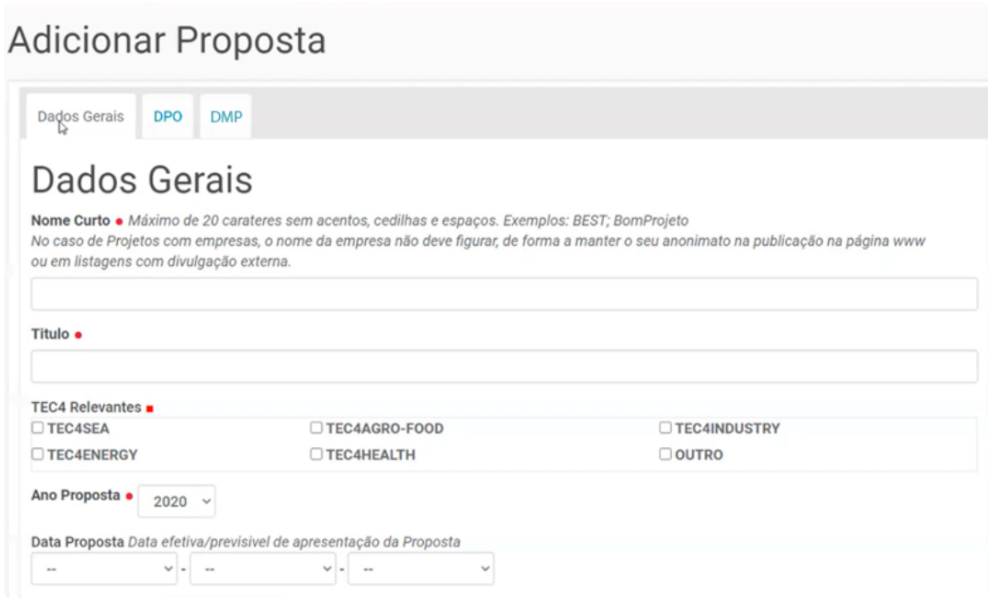

Figure 69: DMP tab on the INESC TEC Intranet during the creation of a project proposal

posal starts in the same way on the INESC TEC Intranet, and then researchers get access to the DMP tab, where they indicate information about DMP creation needs.

The DMP tab includes questions related to DMP creation (Figure 70). Namely, they indicate if the DMP should be submitted during the project proposal submission, a time during which the DMP should be created and submitted, and the intention to deposit datasets on the institutional repository INESC TEC RDM.

In case the DMP has to be submitted during the project proposal submission, the data steward receives an email after proposal creation, contacts the researcher, and helps with the preparation of the plan. Otherwise, when the DMP should be created after the funder approves the proposal, the data steward receives an email only after the project's approval on the INESC TEC Intranet. Then, all information included in the project proposal is transferred to the IRIS platform. This process is similar to the project submission process described above (Chapter 6, Section 6.1).

During the transfer process, the information contained in the INESC TEC Intranet is automatically inserted in the respective fields of the DMP support system and implemented on IRIS (Figure 71, 72). The inclusion of a DMP support system on IRIS can help automate the creation of plans, making them compliant with the

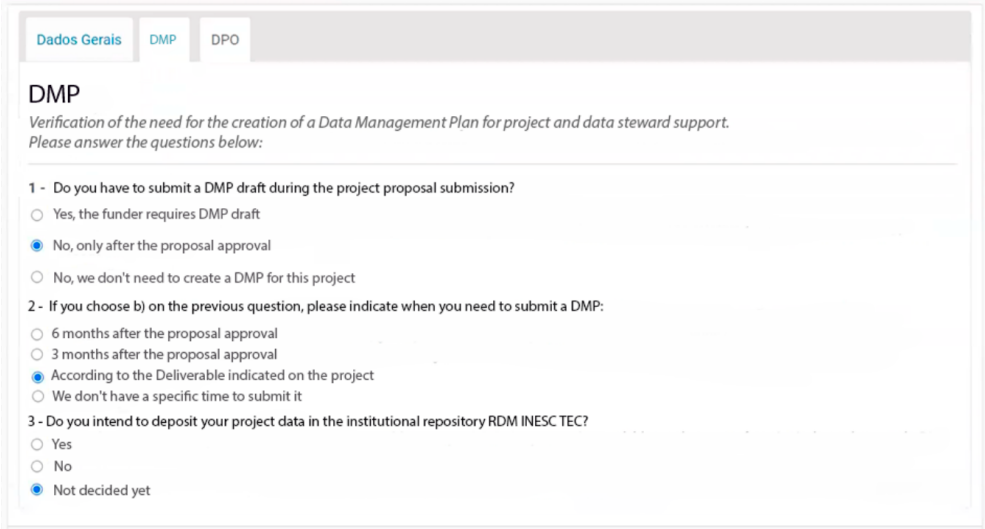

Figure 70: Questions for DMP support request during the creation of a project proposal on the INESC TEC Intranet

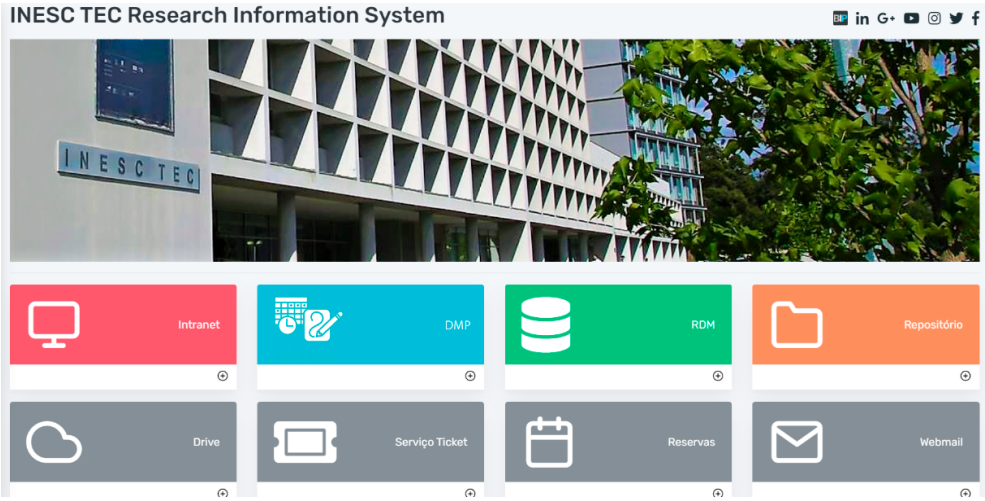

Figure 71: Interface IRIS with DMP support system

maDMP standard. This inclusion also allows having project management, data management, and plan management in one place, facilitating the work of not only the researcher but also the data steward and stakeholders involved on the RDM.

During the transfer process, fields such as the project ID, acronym of the project, abstract, funder, and type of data (sensitive, personal, or private) could be filled out automatically. Moreover, fields such as principal investigator, controller, and DPO could be presented as controlled vocabularies with a drop-down list of terms during the creation of a project proposal in the INESC TEC Intranet (Figure 73). Although these fields are inserted automatically, we suggest making them editable so that errors can be corrected or additional information entered, and existing information improved.

To continue creating the plan, a data steward can add more details to the DMP, indicating all necessary information. Some fields (e.g., metadata standard, techniques for data collection, repository) could also be presented as controlled vocabularies (Figure 73). Some fields, such as the estimated size can automatically notify IT staff by email to prepare the requested size on the server to store the data (Figure 74).

Furthermore, in case there is a request for the DPO or data steward to suggest the revision of the plan by the DPO, it is possible to realize the support request and

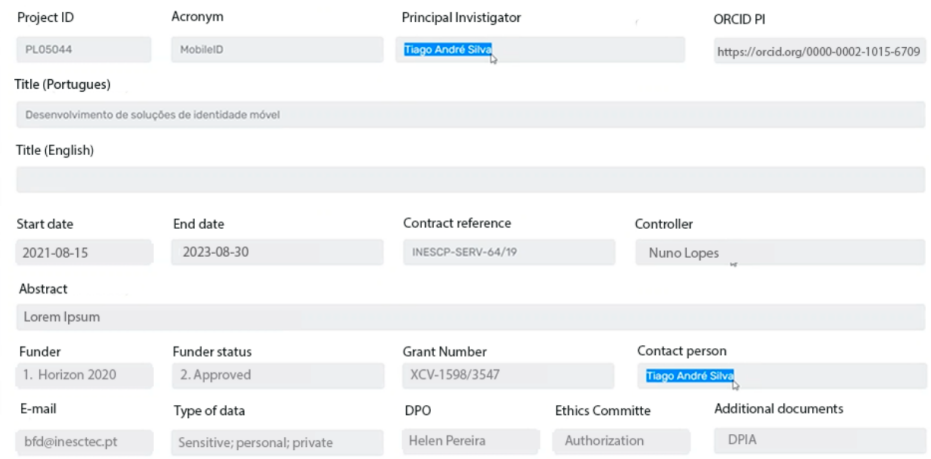

Figure 72: Information automatically inserted on the DMP during transfer process from INESC TEC Intranet to IRIS

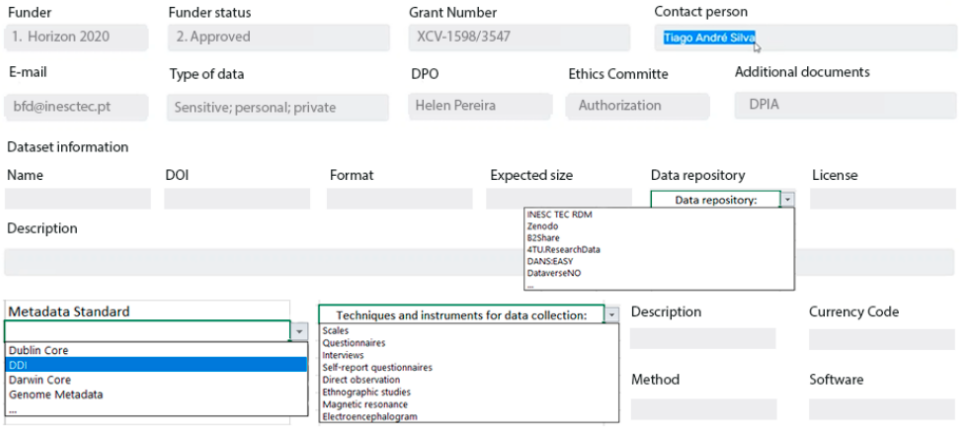

Figure 73: DMP creation interface on IRIS and controlled vocabularies for metadata standards, data repository and instruments for data collection

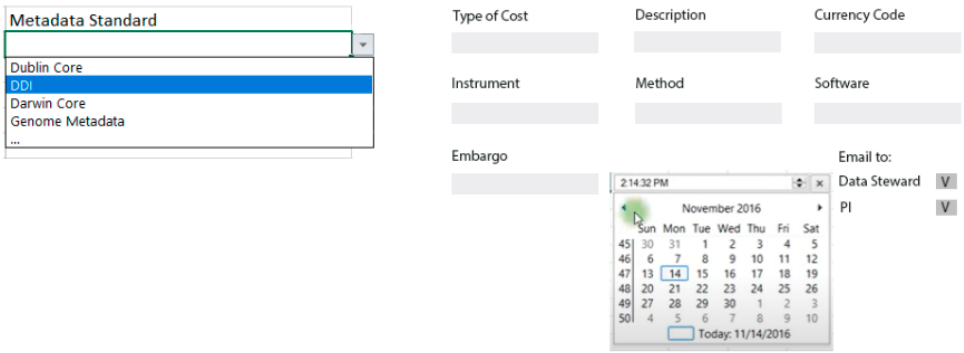

Figure 74: Embargo period and sending email to the data steward or the PI

revision of the DMP in the same place and receive the approval or comments to improve the plan (Figure 75).

If a dataset has some embargo period or restrictions for opening, a data steward could indicate the date when the dataset can be opened, and a few days before that they can automatically receive an email with a warning. In this case, the data

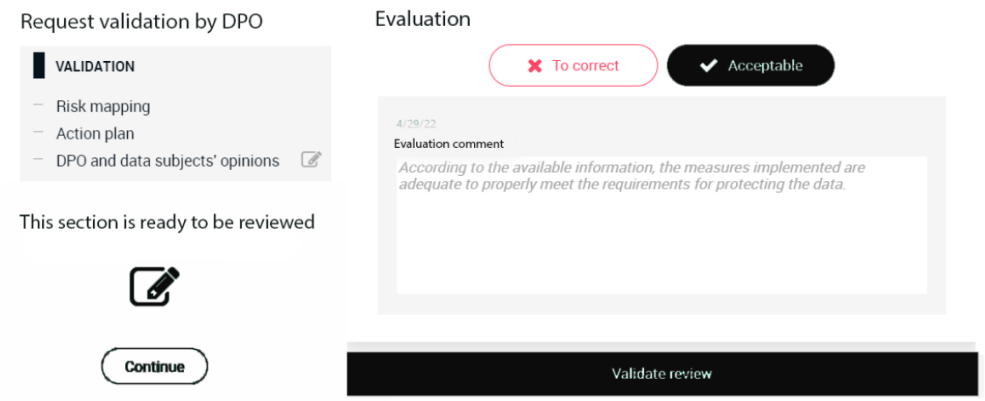

Figure 75: Validation process by the DPO

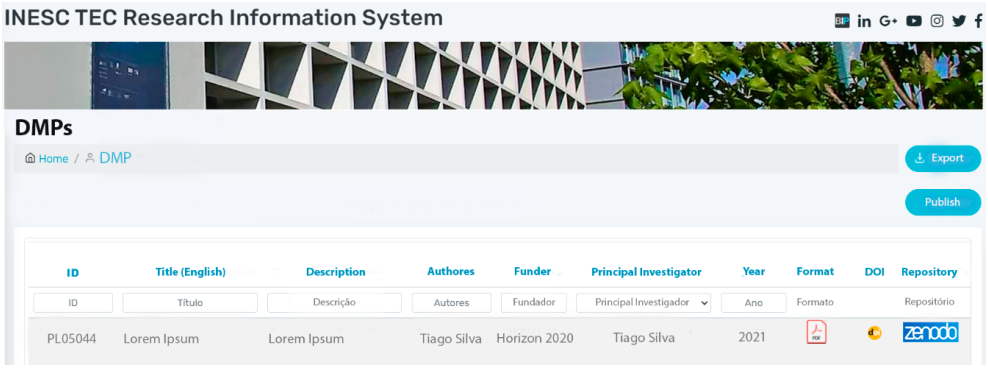

Figure 76: Interface for DMP download and publishing

steward and researchers may decide not to open the dataset until further validation (Figure 74).

When a DMP is created, it can be saved, published or downloaded in different formats, both human-readable and machine-readable DMP (e.g., pdf, word, json, rdf). To publish the plan, a data steward should choose a repository, for example, Zenodo or another trustworthy repository, and send the plan for publication (Figure 76).

In addition, in our DMP support system, we also propose to keep the DMP editable with the tracking of all changes made to it, then saving it as the next version of the plan. The Zenodo repository allows keeping all the versions of the plans (Figure 77).

Finally, we propose the possibility of the data steward scheduling the monitoring session for the DMP with automatic reception of e-mail with warning (Figure 78).

6.4 SUMMARY

Developing a DMP support system involves analyzing different aspects of the institution. It is a complex process that requires effort, time and different types of collaboration. It involves collaboration not only with researchers to analyze their needs but also with institutional stakeholders and departments who are in one way or another involved in project and research management and interested in implementing a DMP support system.

In this chapter, we describe an analysis of the institutional project management processes and internal rules using the example of INESC TEC to understand how

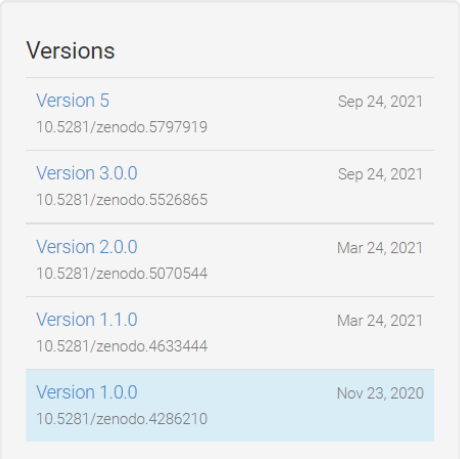

Figure 77: Different versions of the DMP published on Zenodo

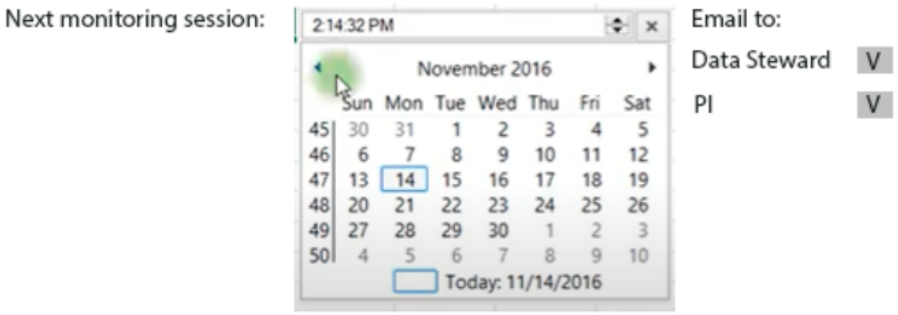

Figure 78: Scheduling for DMP monitor session

the process is organized and how the DMP support system can be implemented in the institutional workflow.

We show that such an analysis should be carried out at the beginning of the development process in order to determine the status of the institutional RDM and DMP workflows. This helps to develop the support system and services, taking into account the institutional processes, researchers and institutional needs and to choose RDM best practices, tools and recommendations regarding DMP in advance.

Moreover, in this chapter, we analyze the DMP workflow in order to make it in line with the maDMP standard, identifying aspects that need to be changed to make our DMP machine-actionable. We propose a new DMP structure, a DMP workflow, and a way to integrate a DMP support system into the institution. We also suggest an interface for this implementation, using controlled vocabularies and automation of the DMP-related mechanisms, such as requests or DMP monitoring schedules.

This analysis help to propose a well-organized institutional DMP support system, which reduces the efforts of the data steward in all DMP-related processes, such as DMP creation, monitoring and publishing. Moreover, our proposed implementation helps to improve institutional workflow by locating all important services and related processes in one place, facilitating project and researcher management, and providing access to all project documentation to all institutional stakeholders and departments.

7 A SYSTEMATIC CONSTRUCTION OF DMP

The collaborative DMP-building method development and its analysis in case studies in Chapter 5 shows how important it is for institutions to provide DMP support and have data stewards who can clarify RDM questions that researchers might have during projects. The previous work served as a basis for the following improvements, showing the importance of having a DMP monitoring process and collaboration between data stewards and other parties as required (e.g. DPO) [145, 146].

To improve the proposed collaborative method and develop a DMP support system as an institutional service, we needed to systematize and test our approach. We also needed to apply this method in different scientific domains to show that it could be used for multiple domains, for any research project and be implemented as a part of the RDM workflow of the institution.

7.1 SAMPLE DEFINITION

To systematize and test our approach, we needed to determine the sample. The projects to be systematically analyzed are a complete sample of research projects at INESC TEC, started in 2020 and 2021. To obtain detailed information about the projects, we contacted the Funding Opportunities Office at INESC TEC and collected the database with **70** projects. This database included a description of the projects approved for funding, where the INESC TEC is either the main organization responsible for the project or one of the project partners (see Figure 79 and dataset "Assessment of metrics for the development of an institutional DMP support system" [131]). Moreover, to select the projects in the sample, we also used 3 filters.

As mentioned above, the first filter was "Year". We selected the projects that started in 2020 and 2021.

The second filter used to select projects for our sample was a "Funder" of the project. We selected projects funded by Horizon 2020[1] and FCT[2]. Horizon 2020 is a European program which supports research projects and provides funding for technological and innovative development from different scientific domains. FCT is the national funding agency that also funds research projects in all scientific domains aiming at science, technology, and innovation development.

The third filter to apply was the date of the end of the project and its completion. Although a DMP can be created at any stage of a project (beginning, middle, or even after completion), for our study, we chose active projects, which were at the beginning or at the middle stages of implementation. Thus, we excluded projects which had been completed before December 31, 2021. With all filters applied to the projects, we received a total of **33** projects that can be seen in Figure 80 and the published dataset "Assessment of metrics for the development of an institutional DMP support system" [131].

1 https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020
2 https://www.fct.pt/index.phtml.pt

nomeCurto	titulo	tipologiaProjeto	tipologiaEspecificaProjeto
CA18130	European Network for Chemical Elemental Analysis by Total Reflection X-Ray Fluorescence	PUE-DIV	Cost Programme
M_NEST_I	Network for Empowering People in Added-Value Manufacturing Systems and Technologies	PUE-H2020	EIT-Manufacturing
SAFECoating	Safety Enhanced Collaborative Robotic Coating Cell	PUE-H2020	Horizon 2020 - Industrial Leadership
TAMI	Transparent Artificial Medical Intelligence	PN-P2020	P2020 em Co-Promoção com CMU - Norte
EIT Jumpstarter	EIT JUMPSARTER Local Trainning - Fase II	PUE-H2020	EIT-Manufacturing
PANDORA	Cyber Defence Platform for Real-time Threat Hunting, Incident Response and Information Sharing	PUE-DIV	European Defence Industrial Development Progr
VitalPROVID	Vital PROVID - sistema de Monitorização de Doentes - Durante e após o COVID-19	PN-P2020	P2020 I&D Empresas - POCI
Sail2020	Interações Espaço-Atmosfera-Oceano na Camada Limite Marinha	OID	Fundo Ambiental
NEWSAT	Development of a compact integrated sensor and satellite for earth observation	PN-P2020	P2020 em Co-Promoção com MIT - NORTE
ATTEST	Advanced Tools Towards cost-efficient decarbonisation of future reliable Energy SysTems	PUE-H2020	Horizon 2020 - Low Carbon Energy
EUniversal	MARKET ENABLING INTERFACE TO UNLOCK FLEXIBILITY SOLUTIONS FOR COST-EFFECTIVE MANAG	PUE-H2020	Horizon 2020 - Low Carbon Energy
EENPortugal	EEN-Portugal 2020-2021	PUE-H2020	Horizon 2020 - INNOVATION
ZeroDefects40	Advanced decision making for zero defects manufacturing	PUE-H2020	EIT-Manufacturing
M_NEST_RIS	Creating cross-European networks of Partners with complementary competences in targeted AVM technologi	PUE-H2020	EIT-Manufacturing
UNEXUP	UNEXUP. UNEXMIN Upscaling	PUE-H2020	EIT-RawMaterials
EIT_M_RIS_Hubs	EIT Manufacturing RIS Hubs	PUE-H2020	EIT-Manufacturing
CPPS101	Cyber-physical production systems	PUE-H2020	EIT-Manufacturing
UP4Legacy	Solutions to transform legacy metal working machines into modern cyber-physical systems	PUE-H2020	EIT-Manufacturing
DigTrafoRIS	Digital Transformation in RIS	PUE-H2020	EIT-Manufacturing
EENINNOVATION	EEN-INNOVATION_JOURNEY_2020-2021	PUE-H2020	Horizon 2020 - INNOVATION
AI_REGIO	Regions and DIHs alliance for AI-driven digital transformation of European Manufacturing SMEs	PUE-H2020	Horizon 2020 - Information and Communication T
QuALTOS	Quality Assurance in Long Term Observation Systems	PN-FCT	FCT com financiamento exclusivo OE
CHARD_IP	CHARD - Apoio à internacionalização da Propriedade Intelectual	OID	Proteção de Direitos de Propriedade Intelectual/Ir
SHIELD	Proteção do Património Cultural contra Pilhagem em Distritos em Perigo	PUE-DIV	European Research Area - NET - H2020
INCAFO	Identificação de Castas através da folha	PN-P2020	P2020 em Co-Promoção - POCI
BATERIAS2030	BATERIAS 2030 - As baterias como elemento central para a sustentabilidade urbana	PN-P2020	P2020 Mobilizador - POCI
PLASMA 4 COVID	Agilização da colheita de plasma convalescente	PN-FCT	RESEARCH 4 COVID-19

Figure 79: Database with the information of the project provided by the Funding Officer

projectId	shortName	title	acronym	beginDate	expectedEnd	centros	inescte	totalInc	nome
20982	QuALTOS	Quality Assurance in Long Term Observation Systems	PN-FCT	01/01/2020	31/12/2022	CRAS	99396	99396	Nuno Cruz
21522	THOR	THOR - Avaliação Torácica Assistida por Computador usando POCUS	PN-FCT	01/03/2021	29/02/2024	CBER	101285	239665	Miguel Coimbra
21532	FronTowns	Pensa em grande sobre as pequenas vilas de fronteira: Alto Alentejo e Alta Extremadura leonesa (séculos XIII – XVI)	PN-FCT	20/03/2021	19/03/2024	CSIG	33905	188159	Leonel Morgado
21563	PassCert	PassCert: Investigação do Impacto de Verificação Formal na Adopção de Software para Segurança de Passwords	PN-FCT	01/02/2021	31/01/2022	HASLAB	28550	69994	José Bacelar Almeida
21571	CAGED	Diagnóstico de Cancro Gástrico Assistido por Computador	PN-FCT	01/03/2021	29/02/2024	CBER	210200	249460	Miguel Coimbra
21589	Connect2Oceans	Conectar os Oceanos Atlântico e Ártico para Decifrar o Impacto das Alterações Climática nas Funções dos Microbiomas Plantónicos	PN-FCT	20/03/2021	19/03/2024	CRAS	48426	249330	Alfredo Martins
21590	MATinMOL	Ondas de matéria em Redes Moiré	PN-FCT	01/03/2021	29/02/2024	CTM	25470	243288	Maria Inês Carvalho
21595	FuturePharma	Cadeia de abastecimento farmacêutica do futuro	PN-FCT	29/03/2021	29/03/2024	CESE	58768	249362	Jorge Pinho de Sousa
21678	BeFresh	On incorporating consumer behaviour into the supply chain planning of fresh products	PN-FCT	01/05/2021	01/05/2024	CEGI	239090	239090	Pedro Amorim
21680	MYTAG	MyTag – Dispositivos de monitorização ambiental universais baseados na análise Inteligente de dispersão ótica amplificada com polímeros de impressão molecular.	PN-FCT	01/01/2022	01/01/2025	CAP	109006	244498	Pedro Jorge
20893	inSITE	inSITE. Insitu ore grading system using LIBS in harsh environments	PUE-H2020	01/01/2020	31/12/2022	CAP,CR	592500	1937146	Pedro Jorge
20938	ATLANTIS	The Atlantic Testing Platform for Maritime Robotics: New Frontiers for Inspection and Maintenance of Offshore Energy Infrastructures	PUE-H2020	01/01/2020	31/12/2022	CRAS	2296250	7046271	Andry Maykol Pinto
20958	ATTEST	Advanced Tools Towards cost-efficient decarbonisation of future reliable Energy SysTems	PUE-H2020	01/03/2020	28/02/2023	CPES	744250	3998659	Filipe Joel Soares
20959	EUniversal	MARKET ENABLING INTERFACE TO UNLOCK FLEXIBILITY SOLUTIONS FOR COST-EFFECTIVE MANAGEMENT OF SMARTER DISTRIBUTION GRIDS	PUE-H2020	01/02/2020	31/07/2023	CPES	807500	7999997	Clara Sofia Gouveia
20971	UNEXUP	UNEXUP. UNEXMIN Upscaling	PUE-H2020	01/01/2020	31/12/2022	CRAS	779059	2992588	José Miguel Almeida
21355	AI_REGIO	Regions and DIHs alliance for AI-driven digital transformation of European Manufacturing SMEs	PUE-H2020	01/10/2020	30/09/2023	CESE,CI	156250	9200080	César Toscano
21379	BetterFactory	Grow your manufacturing business' — 'Better Factory'	PUE-H2020	01/10/2020	30/09/2024	CESE,CI	259250	8000342	Germano Veiga
21384	OneNet	One Network for Europe	PUE-H2020	01/10/2020	30/09/2023	CPES	196875	2,2E+07	Alexandre Lucas
21404	TRUSTAI	TRUST-AI: Transparent, Reliable and Unbiased Smart Tool for AI	PUE-H2020	01/10/2020	30/09/2024	CEGI	898750	3995418	Gonçalo Reis Figueira
21407	NOVATERRA	PESTICIDES, TOWARDS SUSTAINABLE MEDITERRANEAN VINEYARDS AND OLIVE GROVES	PUE-H2020	01/10/2020	30/09/2024	CRIIS	341187	4999908	Filipe Neves Santos
21411	PRySM	PRecision Sprayer Ground Robot	PUE-H2020	01/05/2020	10/01/2022	CRIIS	93250	190750	Filipe Neves Santos
21434	SCORPION	Cost effective robots for smart precision spraying	PUE-H2020	01/01/2021	31/12/2023	CITE,CR	413987	2270606	Filipe Neves Santos
21445	HumanE-AI-Net	HumanE AI Network	PUE-H2020	01/09/2020	31/08/2023	LIAAD	110000	1,2E+07	João Gama
21479	MARI4_YARD	User-centric solutions for a flexible and modular manufacturing in small and medium-sized shipyards	PUE-H2020	01/12/2020	30/11/2024	CRIIS	492375	4998825	Germano Veiga
21547	CircThread	Building the Digital Thread for Circular Economy Product, Resource & Service Management	PUE-H2020	01/06/2021	31/05/2025	CESE,HI	328250	7994956	António Lucas Soares
21560	EIT_RIS_Hubs_20	EIT Manufacturing RIS Hubs 2021	PUE-H2020	01/01/2021	31/12/2022	CESE	111350	606500	Ana Cristina Barros
21569	Inno4Vac	A cloud-based systems-immunology platform for reliable predictions of vaccine efficacy	PUE-H2020	01/09/2021	28/02/2027	CSIG	298750	2000000	Artur Rocha
21588	RISC2	A network for supporting the coordination of High-Performance Computing research between Europe and Latin America	PUE-H2020	01/01/2021	30/06/2023	HASLAB	26250	499911	António Luís Sousa
21592	PAFSE	Partnerships for science education	PUE-H2020	01/09/2021	01/05/2024	CSIG	175000	1425000	Hugo Paredes
21613	EUSCORES	EUropean - Scalable and Complementary Offshore Renewable Energy Sources	PUE-H2020	01/09/2021	31/08/2025	CAP,CE	2221375	3,5E+07	Bernardo Silva
21620	MAGPIE	sMArt Green Ports as Integrated Efficient multimodal hubs	PUE-H2020	01/10/2021	01/10/2026	CESE,CI	917937	2,5E+07	David Emanuel Rua
21627	FIRE_RES	Innovative Technologies and Socio-Ecological-Economic Solutions for FIRE RESilient Territories in Europe	PUE-H2020	01/12/2021	01/12/2025	CBER,CI	781375	2,2E+07	Alexandra Xavier
21632	FIRELOGUE	Cross-sector dialogue for Wildfire Risk Management	PUE-H2020	01/12/2021	01/12/2025	CRAS	120312	3259371	Hugo Miguel Silva

Figure 80: Summary of the projects in our sample

7.2 PRE-ANALYSIS OF THE SAMPLE

Before contacting researchers for collaboration, we analyzed our sample, gathered information about the projects, such as the context and name of the PI, and made conjectures about the scientific domain and type of data collected during the project (public, sensitive, private, personal). We also prepared all necessary documents for collaboration, such as a guide for the meeting (see Appendix B), the text of the email for contact (see Appendix C), and a confidentiality agreement proposed to researchers if they agree to participate in our study (see Appendix D).

The preliminary assignment about the scientific domain and data type allowed the data steward to prepare for the meeting in advance, verifying specific aspects of the domain, as well as to be more focused on certain issues depending on the data types. For example, projects collecting personal data usually require the creation of additional documents, such as informed consent, Data Protection Impact Assessment, or authorization from the Ethics Committee. Moreover, some of them might require collaboration with other stakeholders, such as the DPO.

The preliminary assignment of a scientific domain also helps the data steward to prepare for a meeting with more specific questions for such projects and verify the existence of the controlled vocabularies for the domain. It also helps to try

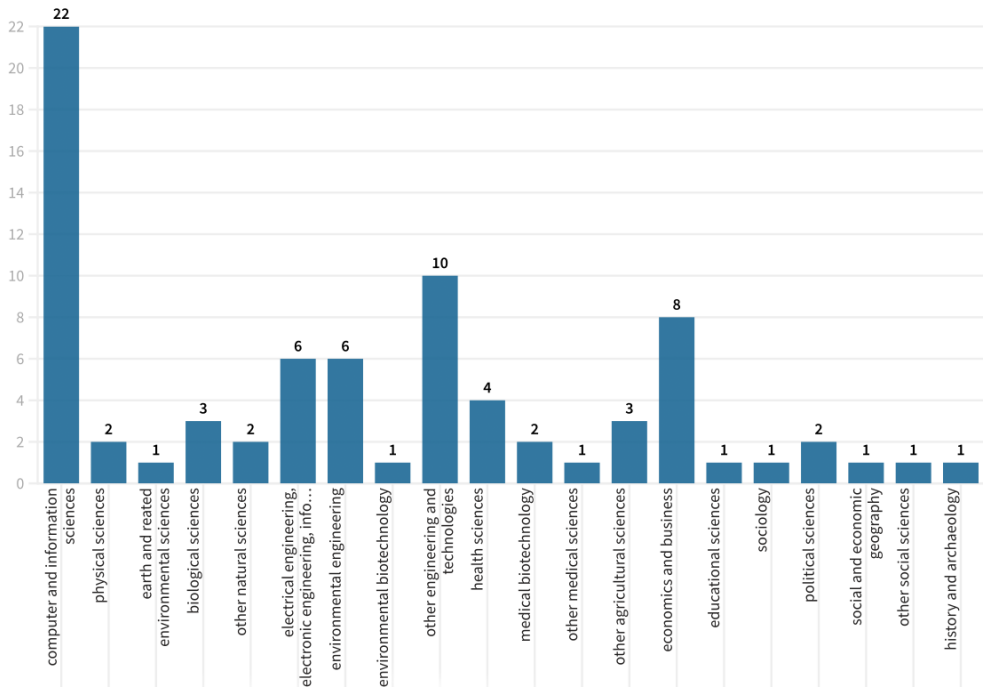

Figure 81: Distribution of the scientific domains in the sample for study, according to FOS classification and according to the data steward preliminary assignment

to identify domain-specific terms that can be used when creating Domain Data Protocols, which in turn can be used when creating DMPs.

We also identified, where possible, the countries of the partners involved in the project, as the transfer of personal data to some countries, for example, may require the creation of additional agreements. In other words, the data steward, knowing which countries are included in the project, can also ask more specific questions on this aspect.

To standardize scientific domains in our sample, we based our preliminary assignment on the Field of Science and Technology (FOS) classification in the Frascati Manual [248]. There are six principal fields: Natural Sciences, Engineering and Technology, Medical and Health Sciences, Agricultural Sciences, Social sciences and Humanities, and forty-two related sub-fields.

To apply the FOS classification, we analyzed the context of the projects from their web pages, assuming the projects' scientific domains. Moreover, we made a preliminary assignment about the types of data collected during the project, also based on the analysis of the project context and on the professional experience of the data steward, who had already supported projects of different domains in RDM and DMP issues.

Figure 81 represents the domains to which the projects belong according to our preliminary assignment. Some projects may belong to more than one domain, being diverse and complex. As a result, most of the projects (**22**) belong to the "computer and information sciences" domain; the domains "earth and related environmental sciences", "environmental biotechnology", "other medical sciences", "educational sciences", "sociology", "social and economic geography", "other social sciences" and "history and archaeology" include only **1** project each (see Appendix E).

During the analysis of the project context, we also made a preliminary assignment about the type of data the project might collect so that the data steward could ask for more details during the interview about any specific situations related to the identified types of data. In this case, the interview would include more attention to complex questions related to additional documents and resources related to sensitive, personal, or private data. In Figure 82, we present the distribution of

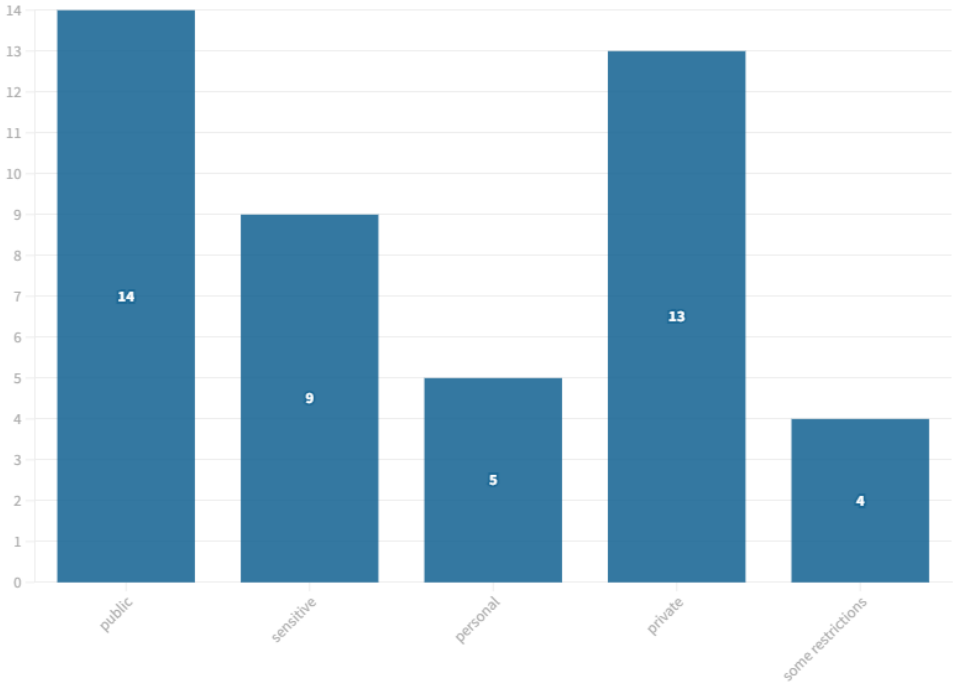

Figure 82: Distribution of the type of the data according to the preliminary assignment of the data steward during pre-analysis

the different types of data to be collected according to our preliminary assignment based on the pre-analysis as well as the professional experience of the data steward. As a result, we have: **14** projects that collect public data, **9** - sensitive data, **5** - personal data, **13** - private data, and **4** - public data with some restrictions such as embargo period.

In addition, we contacted the DPO at INESC TEC to check which projects on the sample list have already requested support from the DPO team. This might be related to sensitive, personal or private data, support in creating a DMP, clarification of general issues, or support in creating additional documents, among other things. For this purpose, we sent the DPO our list of projects and asked them to indicate the projects that had already contacted the DPO team and what exactly the DPO team had supported them on. As a result, we have in **33** projects: **4** projects have requested support to create DMP in the proposal phase, **8** projects have requested DPO support related to other issues related to sensitive and personal data, cookie policy and terms of use, data protection and ethics issues, and revision of the documents, such as partner agreements, surveys, and informed consent (Figure 83).

This type of analysis allowed us to identify projects in which researchers had already thought about RDM issues, which in turn allowed us to check what methods and tools they used to create DMPs or other relevant documents. It also allowed us to make suggestions to researchers to improve already existing documents and monitor them.

As a result of the project selection and pre-analysis, we created an Excel file structure, "The project sample", which is published on the INESC TEC RDM repository together with other files as our dataset "Assessment of metrics for the development of an institutional DMP support system" [131], with the following information: project id, shortName, project title, funder, begin date, expected end, the entity in charge of the project, budget, name of the person in charge, a brief description of the project, DMP support according to the DPO answer, Data Protection support according to the DPO answer, scientific domain (our preliminary assignment), scientific domain according to the researchers' answers, experience in data management, type of data (our preliminary assignment), the number of meetings, dates related

Figure 83: Distribution of the projects that requested support from the DPO team

to the DMP creation, monitoring or improvements sessions, DMP status (in case a DMP is published, the DMP link/DOI and date of the publication), and specific terms defining during creation of the DMP for Domain Data Protocols (controlled vocabulary). Moreover, we also indicated what countries entered into the projects and who is the coordinator. This type of analysis can help us study cases in which a DMP may be created by other partners and offer support for publishing, improving and monitoring the DMP, if needed. The file structure and information in it were updated as the analysis proceeded, adding the necessary columns.

This preparation step and analysis of the information about the projects helped the data steward to be familiar with the projects, classify projects according to Frascati classification (see Appendix E), define the evaluation metrics presented in Section 7.4 of the collaborative method that we describe below, think about specific points and difficulties that researchers could have during the DMP creation and be more attentive to projects in which DMP was already created by partners.

7.3 SCENARIOS OF THE DMP PROCESSES

Another preparatory task focused on describing the possible scenarios related to the creation, improvement, publication, or monitoring of DMPs that would help any data steward at an institution to facilitate decision-making when supporting researchers. According to the literature [99], the creation of different scenarios helps to detail the tasks and actions that should be performed to achieve certain goals. Scenarios can be seen as a set of actions validated by conditions. Thus, they are useful when designing systems, their processes, and procedures, giving a general idea of the actions to be performed. Based on our previous work [145], we created **2 scenarios** (Figure 84, Figure 86) to support researchers and detailed the actions and tasks to be performed in each one. Despite most of the projects from the sample are contacted by the data steward, we would suppose that there may also be projects in our sample that themselves request support from the data steward or have already requested it before. Although both scenarios are similar, there are slight differences which we described below. In Figure 84, the common part of both scenarios is highlighted.

In the first scenario the data steward contacts the researchers to verify if they need DMP-related support (see Figure 84). In this case, in the beginning, the data steward verifies whether the researchers **have already created a DMP (Figure 84, condition 1)**, and, depending on the answer, decides what to do next. There are

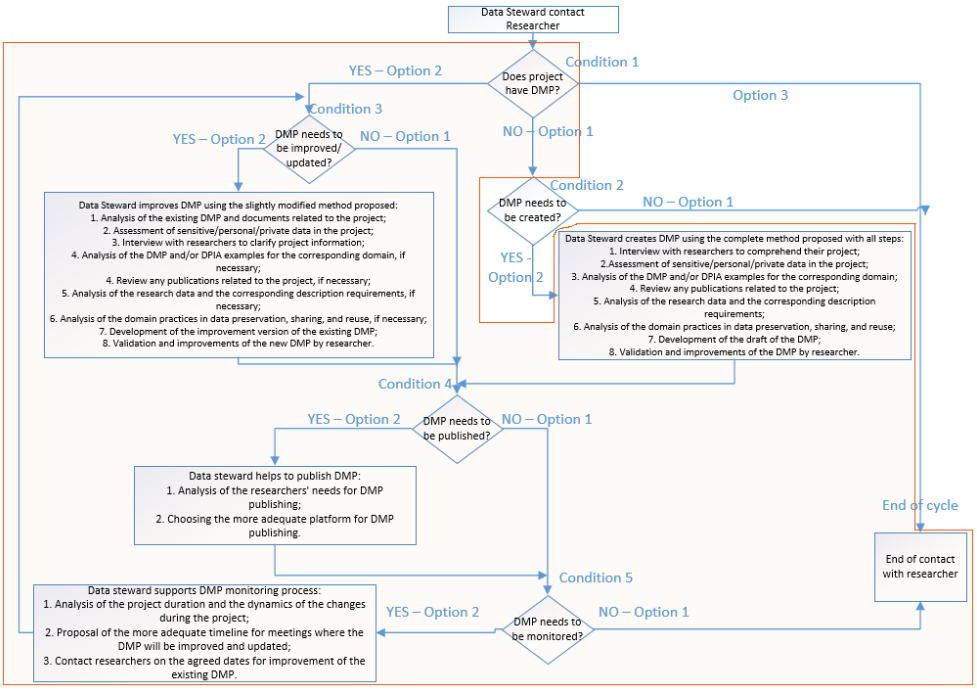

Figure 84: Scenario 1. Data Steward contacts Researcher (highlighting is common to both scenarios)

three options: the researchers do not have a DMP **(Figure 84, condition 1, option 1)**, the researchers already have a DMP **(Figure 84, condition 1, option 2)**, and the researchers do not want to answer this question and do not want any collaboration or support **(Figure 84, condition 1, option 3)**.

If the researchers **do not have a DMP (Figure 84, condition 1, option 1)**, the data steward verifies **whether the researchers need to create a DMP (Figure 84, condition 2)**. There can be two options here: researchers want to create a plan **(Figure 84, condition 2, option 2)** or they do not **(Figure 84, condition 2, option 1)**. If the researchers do not want to create a DMP, the support closes **(Figure 84, end of cycle)**. If they want to create a plan, the data steward initiates the creation process according to the existing collaborative DMP-building method, following all its steps.

The first step is to interview the researchers to understand their project (Step 1 in Figure 85); the second one is to determine if the project is ongoing contain any personal, sensitive, or private data (Figure 85, Step 2); the third step is to analyze the DMP examples and, in cases when DPIA is required, analyze DPIA examples for the corresponding domain (Figure 85, Step 3); the fourth step is to review researchers publications related to the project (Figure 85, Step 4); fifth, an analysis of the project's research data and corresponding description requirements (Figure 85, Step 5); next, an analysis of the domain practices in data preservation, sharing, and reuse (Figure 85, Step 6); then the creation of the DMP draft (Figure 85, Step 7); and its presentation for review and content improvement by researchers (Figure 85, Step 8). In other words, in this scenario, the data steward uses the complete proposed collaborative method with all the steps described in Chapter 5 and proceeds to analyze the next condition related to the **publication of the DMP (Figure 84, condition 4)**.

If researchers have a DMP **(Figure 84, condition 1, option 2)**, the data steward immediately proceeds to check the next condition related to the **DMP improving/updating (Figure 84, condition 3)**.

In the third option (Figure 84, condition 1, option 3) researchers do not want to respond. Since we cannot analyze the reasons for the rejection of our collaboration

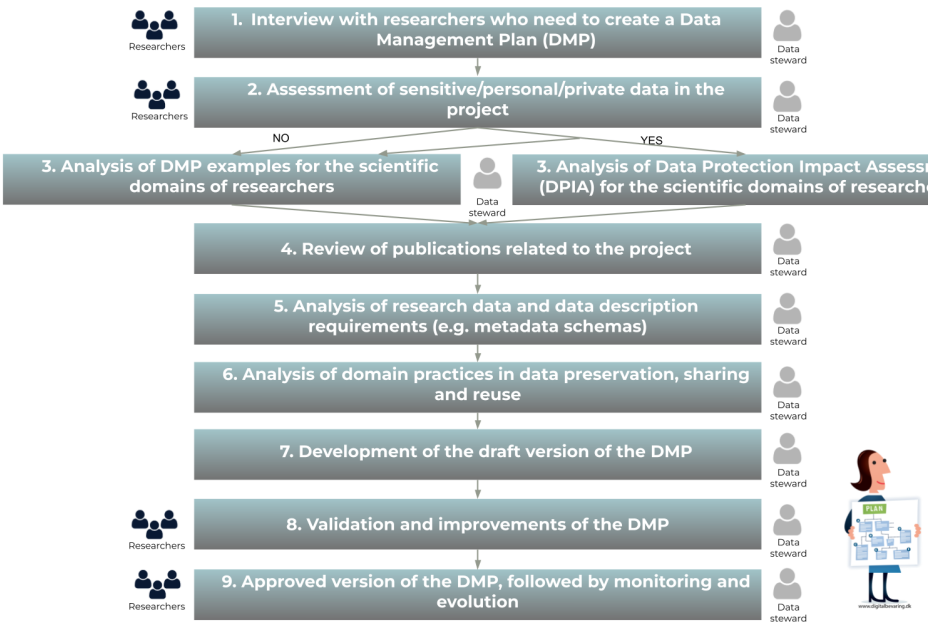

Figure 85: The collaborative DMP-building method (i.e. Figure 52)

and provide justification, we will simply leave these cases unsupported **(Figure 84, end of cycle).**

When the DMP has already been created **(Figure 84, condition 1, option 2)**, the data steward checks to see if the researchers need to improve/update their plan **(Figure 84, condition 3)**. There can be **two possibilities here**: they do not need to improve/update it **(Figure 84, condition 3, option 1)** and it needs some improvements/updates **(Figure 84, condition 3, option 2)**. If the DMP does not need improvement/update, the data steward moves on to check the next condition related to the publication of the DMP **(Figure 84, condition 4)**.

If the DMP needs to be improved, then the data steward follows the collaborative DMP-building method and, after the first meeting with the researchers proceeds to analyze the existing DMP and the documents related to the project, such as consortium agreements. The data steward analyzes the context of the DMP, the presence of sensitive, personal or private data, and relevant documents such as the DPIA or the ethics committee authorization. The data steward also organizes a meeting where more details about the changes that have occurred in the project since the creation of the DMP, changes related to the organization of the new datasets, and changes in the license and repository selection are asked. Moreover, in this meeting, the data steward clarifies already collected information related to the RDM in the project. After that, if necessary, the data steward follows the remaining steps of the collaborative method, proposes an improved version of the existing DMP, and proceeds to analyze the next condition related to the **publication of the DMP (Figure 84, condition 4).**

If the researchers do not want to **publish their plan (Figure 84, condition 4, option 1)**, the data steward moves on to analyze condition number 5 related to the **monitoring process (Figure 84, condition 5)**. If they choose the option of publishing a DMP **(Figure 84, condition 4, option 2)**, the data steward helps select a more appropriate platform according to the needs of the researchers by analyzing the publishing goals. These can be related to the dissemination of project information, the need to have a machine-readable DMP, linking the plan to project data, etc. For example, Zenodo can be chosen, because there the plan can be assigned a DOI, the DMP can be open for dissemination, even if the project data cannot be opened, the DMP can serve as one of the project outputs and can be updated and saved with all versions, thus tracking the changes occurring during the project.

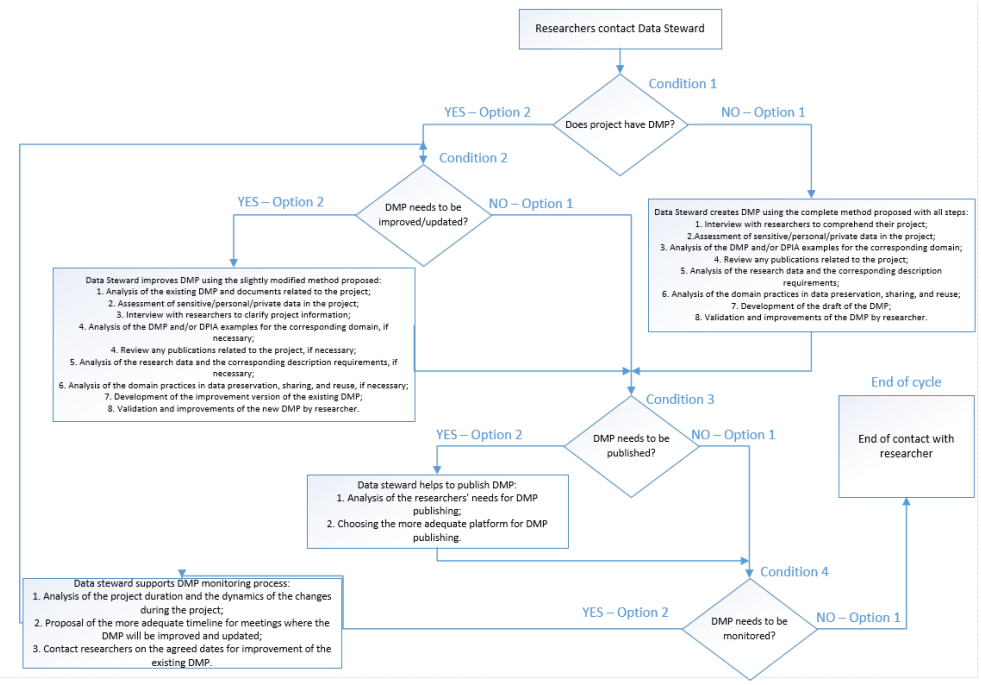

Figure 86: Scenario 2. Researcher contacts Data Steward

The next condition that the data steward checks is whether DMP monitoring is necessary **(Figure 84, condition 6)** or not. If researchers do not need to monitor their plans **(Figure 84, condition 6, option 1)**, support is closed (**Figure 84, end of cycle)**. For example, this might happen at the end of the project. In contrast, the data steward analyzes the duration of the project, the dynamics of change during the project, proposes a more adequate schedule of meetings where the DMP will be improved and updated, adds the DMP to the database as a formal document for further monitoring, contacts the researchers on the agreed dates for improvement of the existing DMP **(Figure 84, condition 3)**, and continues support until project completion **(Figure 84, end of cycle)**.

The second scenario, where researchers contact the data steward regarding the need to create, publish, improve/update, or monitor a DMP **(see Figure 86)**, is similar to the first scenario **(Figure 84)** with some changes in the process. First of all, the data steward checks whether the researchers **have created a DMP (Figure 86, condition 1)**. If the researchers do not have a DMP **(Figure 86, condition 1, option 1)**, then the data steward uses the complete existing collaborative DMP-building method, following all steps **(see Figure 85)**. This DMP creation process is similar to that described in the first scenario, where the data steward contacts researchers. It includes the same steps of the collaborative DMP-building method: interviews with researchers, analysis of the existence of the sensitive, personal or private data, analysis of DMP and DPIA examples, review of publications, research data, description requirements, RDM domain practices, then the first draft DMP is created, validated and improved by researchers. Once the DMP is created, the data steward proceeds to verify the next condition related to the **publication of the DMP (Figure 86, condition 3).** In the second scenario, there is no option 3 of the first scenario, where researchers do not want to respond or do not want any collaboration or support. Moreover, there is no condition that verifies if the DMP needs to be created because the researchers only contact the data steward when necessary to create the DMP. This saves the data steward time, as well as the data stewards' workload in general.

If the researchers already have a DMP **(Figure 86, condition 1, option 2)**, the situation is similar to the first scenario where the data steward moves on to the next condition related to **improving/updating the DMP (Figure 86, condition 2)**.

To improve the DMP **(Figure 86, condition 2, option 2)**, the data steward should perform different steps of the collaborative DMP-building method (Figure 85), which are also identical to the steps specified in the same "DMP improving/updating" condition in the first scenario **(Figure 84)**. These steps include the analysis of the existing DMP and documents, the existence of sensitive, personal or private data, related documents such as DPIA, and the remaining steps until the researcher's DMP is improved, making it ready for **publication (Figure 86, condition 3)**. If the DMP does not require improvements/updates, the data steward moves on to the next condition related to the **publication of the DMP (Figure 86, condition 3)**.

As in the first scenario, checking the DMP publication condition has two options: researchers want to publish or not. In both options, the condition that follows concerns DMP monitoring **(Figure 86, condition 4)**. If the researchers do not want to publish **(Figure 86, condition 3, option 1)**, the data steward immediately proceeds to check the DMP monitoring condition. In the other option **(Figure 86, condition 3, option 2)**, tasks are performed as in the first scenario, namely, analyzing research needs in relation to DMP publication and selecting the most appropriate repository for publication.

The next option is **DMP monitoring**. If DMP monitoring is not required **(Figure 86, condition 4, option 1)**, support closes **(Figure 86, end of cycle)**. As in the first scenario, the data steward analyzes the duration of the project, the dynamics of change, proposes an adequate schedule of monitoring sessions, and contacts the researchers on the agreed dates to update the DMP. Finally, the data steward returns to the improvements/updates and publication conditions and repeats the same cycle until it ends **(Figure 86, end of cycle)**, when support is no longer needed.

These two scenarios are very similar, but there are peculiarities that depend on who is contacting whom. In our opinion, when the data steward contacts the researchers, they are more likely to refuse support than when the researchers contact the data steward. This may be because when researchers contact the data steward, they demonstrate their willingness, interest, needs, and motivation to create, improve, publish, or monitor the DMP.

7.4 METRICS FOR DMP SUPPORT EVALUATION

To improve the collaborative DMP-building method, systematize it, and implement it as a DMP support system in the institution, we had to test and evaluate it in practice. For this purpose, we established metrics, which included both quantitative and qualitative evaluation.

Quantitative metrics (Quant) are based on previous work and case studies, and result from the answers to the following questions:

– ***Quant1 (Number of acceptance (scenario 1))***: How many projects have accepted (scenario 1) support in DMP creation/publishing/improving/monitoring?

– ***Quant2 (Number of requests (scenario 2))***: How many projects required (scenario 2) for support to create/publish/improve/monitor a DMP?

– ***Quant3 (Sensitive, personal, private data issues)***: How many projects have received help from a data steward for questions related to sensitive, personal, or private data?

– ***Quant4 (Repository issues)***: How many projects have received help from a data steward in choosing a more appropriate repository for publishing data and DMP?

– ***Quant5 (Costs issues)***: How many projects have received help from the data steward in technical aspects such as early purchase of disks for data preservation, storage, and other technology-related costs?

– ***Quant6 (Additional Documents)***: How many projects have required the creation of additional documents, such as DPIA, and informed consent, among others?

– ***Quant7 (Collaboration with other stakeholders)***: How many projects require collaboration with other stakeholders, such as DPO, Ethics Committee, etc?

– ***Quant8 (Ethical and property rights issues)***: How many projects have received help from the data steward with ethical, legal, copyright, Intellectual Property Rights and license issues?

– ***Quant9 (Time)***: How long did it take to create the DMP as a whole?

Along with quantitative metrics, we have identified ***qualitative metrics (Qual)***, which also can help evaluate the collaborative DMP-building method, DMP support in general, and identify points for improving the RDM workflow. For this purpose, we based the metrics on the following questions:

– ***Qual1 (Motivation)***: How enthusiastic and motivated were the researchers from the projects we supported?

– ***Qual2 (Effort)***: Based on the "Motivation" metric, we can also analyze how much effort researchers put into DMP creation without data steward support?

– ***Qual3 (Researchers Satisfaction Score)***: How satisfied are researchers with the DMP support?

– ***Qual4 (Rejection)***: What is the reason for rejection of DMP support?

– ***Qual5 (Impact)***: What was the impact of DMP creation on a project in the form of qualitative metrics based on researchers' opinions?

All of these metrics help us highlight important points in the collaborative DMP-building method, both positive and negative aspects that could be improved. We have also added some of these metrics and results to our Excel "The project sample" that is included in our dataset, which is published on the INESC TEC RDM repository [131].

7.5 SAMPLE DESCRIPTION

Work on projects in the sample started with sending emails to the PI projects. We divided the projects into several subsets, sending 10-12 emails at a time, because if all the projects agreed to collaborate at the same time, the data steward would not be able to give quality support related to the DMP. Creating, monitoring, and analyzing an already created plan or improving the DMP are quite time-consuming processes, which require analysis of documentation, familiarization with the project, understanding of the context, several meetings and interviews, analysis of collected data, and creation of individual plans for each project, taking into account their specifics aspects.

The data steward received immediate responses to some emails, and others were not responded to by PIs, and there could be various reasons for this, which are analyzed in more detail in the following chapter. However, the lack of a response may be due: the email was in spam; along with the huge number of emails received in a day, it could get lost; the PI did not understand our proposal correctly or did not have the time or motivation to create a plan; or they could simply not be interested in collaboration. One of our principal goals of this work is to suggest and implement a well-established DMP support system, and to do this, we also had to analyze the motives for the rejection of collaboration. Therefore, the data steward looked for different ways to communicate with the PI in order to identify these reasons for further analysis.

Thus, when the data steward did not receive any response, the email was sent again, then the data steward tried to reach out through other researchers and professors who had already collaborated with the data steward and the PI. This could be either face-to-face contact or contact through third parties.

Summing up, we have **14** projects from **33** that received data steward support in DMP creation, monitoring, or improvement; **15** projects where the data steward established contact but DMP support did not advanced or the data steward is still waiting for the contact of the PI, and **4** projects where the data steward could not

Table 19: Status of the collaboration of 33 projects in our sample

Project title	Collaboration
Qualtos: Quality Assurance in Long Term Observation Systems	Yes
FronTowns: Think big on small frontier towns: Alto Alentejo and Alta Extremadura leonesa (13th - 16th centuries)	Yes
LifeSkillsVR: Life Skills for Employment in COVID-19 Era through VR Innovation	Yes
FuturePharm: Pharmaceutical supply chain of the future	Yes
ATTEST: Advanced Tools Towards cost-efficient decarbonisation of future reliable Energy SysTems	Yes
EUniversal: Market enabling interface to unlock flexible solutions for cost-effective management of smarter distribution grids	Yes
TRUST-AI: Transparent, Reliable and Unbiased Smart Tool for AI	Yes
NOVATERRA: Integrated novel strategies for reducing the use and impact of pesticides, towards sustainable mediterranean vineyards and olive groves	Yes
PRySM: PRecision Sprayer Ground Robot	Yes
SCORPION: Cost effective robots for smart precision spraying	Yes
CircThread: Building the Digital Thread for Circular Economy Product, Resource & Service Management	Yes
EIT_RIS_Hubs_2021: EIT Manufacturing RIS Hubs 2021	Yes
Inno4Vac: A cloud-based systems-immunology platform for reliable predictions of vaccine efficacy	Yes
PAFSE: Partnerships for science education	Yes
VR2CARE: Multiuser immersive solution for safe group training	Yes
FIRE_RES: Innovative Technologies and Socio-Ecological-Economic Solutions for FIRE RESilient Territories in Europe	Yes
SCRelProg: Self and Co-regulation in e-Learning of Computer Programming:	Yes
SAIL: Space-Atmosphere-Ocean Interactions in the marine boundary Layer	Yes
Future Cities: Urbansense	Yes
aMILE: Application of text mining to clinical reports of patients with acute myeloid leukemia	Yes
Genome Editing: Facilitated Genome Editing as Responsible Research and Innovation	Yes
VigilHate: "Vigilant Citizens Against Hate": How to counter bystander apathy and increase citizens' commitment against online hate speech?	Yes
N-DeciMa: Neurophysiological bases of decision-making processes	Yes
Submerse: SUBMarine cablEs for ReSearch and Exploration	Yes
THOR: Computer Assisted Thoracic Assessment using POCUS	Stand-by
CAGED: Computer Assisted Gastric Cancer Diagnosis	Stand-by
Connect2Oceans: Connecting Atlantic and Arctic Oceans to Decipher Climate Change Impact on Plankton Microbiome Functions	Stand-by
ATLANTIS: The Atlantic Testing Platform for Maritime Robotics: New Frontiers for Inspection and Maintenance of Offshore Energy Infrastructures	Stand-by
FIRELOGUE: Cross-sector dialogue for Wildfire Risk Management	Stand-by
MATinMOL: Matter Waves in Moiré Lattices	Stand-by
BeFresh: On incorporating consumer behaviour into the supply chain planning of fresh products	Stand-by
MYTAG: Universal environmental monitoring devices based on Intelligent optical scattering analysis amplified with molecularly imprinted polymers	Stand-by
inSITE: Insitu ore grading system using LIBS in harsh environments	Stand-by
UNEXUP: UNEXMIN Upscaling	Stand-by
AI_REGIO: Regions and DIHs alliance for AI-driven digital transformation of European Manufacturing SMEs	Stand-by
OneNet: One Network for Europe	Stand-by
HumanE-AI-Net: HumanE AI Network	Stand-by
EUSCORES: EUropean - Scalable and Complementary Offshore Renewable Energy Sources	Stand-by
MAGPIE: sMArt Green Ports as Integrated Efficient multimodal hubs	Stand-by
PassCert - Exploring the Impact of Formal Verification on the Adoption of Password Security Software	No
BetterFactory - Grow your manufacturing business	No
MARI4_YARD - User-centric solutions for a flexible and modular manufacturing in small and medium-sized shipyards	No
RISC2 - A network for supporting the coordination of High-Performance Computing research between Europe and Latin America	No

get any response in all possible ways. Table 19 shows the summary of these projects (detailed information can be seen in a dataset [131]).

7.5.1 Projects with data steward support for DMP

The collaboration with projects in our sample is based on the DMP-building method, considering the projects' needs and highlighting different benefits of DMP creation. The published DMP was seen as one of the benefits of the collaboration, so the data steward was focused on finding a way to ensure that the collaboration would result in a detailed and high-quality document ready for publication or already published. The DMP-light was proposed by the data steward as one of the variants to have the plan ready to publish. This is a short version of the DMP, including a summary of important aspects of research data management in the project, based on already existing DMPs which cannot be published for one reason or another. The proposal of the DMP-light creation could be seen as one of the improvements to our collaborative DMP-building method and the step of verification of its necessity for developing a better-organized DMP support system. The Zenodo platform was chosen as the principal for DMP publishing because it assigns the DOI to each DMP, creates a citation, and promotes the plan as publication, increasing the value of the project and research. The collaboration with the projects took place between November 2021 and June 2022. Now we describe each project in more detail.

Qualtos - Quality Assurance in Long Term Observation Systems

The project, coordinated by INESC TEC, started in 2020 with FCT funding and is related to the calibration of sensors for marine observatories[3]. In this case, the collaboration followed the first scenario where the data steward contacted the researchers to analyze their needs in the DMP creation by sending an email to the project PI. The creation of the DMP took place with their direct involvement and

3 https://www.inesctec.pt/pt/projetos/qualtos

with the involvement of a representative of the Israeli partner. Much of the cooperation was done through e-mail correspondence. During this collaboration, the data steward found out that the PI had no experience of DMP creation or RDM in general and had no contact with the DPO. To create the DMP, the data steward followed the collaborative DMP-building method, which helped to clarify issues related to RDM in general, identify the data collected/created during the project, and verify the repository for data publication. Although they already had a repository dedicated to data collected from sensors, they needed another one for other types of data, and the data steward helped them with. Assistance with other specific points, such as cost issues, creation of additional documents, ethical and property issues, and collaboration with other university departments, were not required. In addition, in this case our preliminary assignment related to the scientific domain was correct: according to the FOS, this project is classified as "other engineering and technologies". As for our preliminary assignment about the existence of private data in this project, this was due to our experience in dealing with data from maritime projects. In some cases, data collected on the sea are seen as private and cannot be opened without an additional agreement. In this case, our project did not contain any private data and all data could be open without any restrictions. Four months of collaboration resulted in the creation of the first draft of the DMP, which is currently under revision.

FronTowns - Think big on small frontier towns: Alto Alentejo and Alta Extremadura leonesa (13th - 16th centuries)

The project started in 2021, led by University NOVA Lisbon with FCT funding. It relates to the historical analysis of the border towns of Portugal and Spain and the creation of a 3D animation model for visualization[4]. Following the first scenario, the data steward sent an email to the PI offering support on DMP creation and the PI referred us to two researchers from Universidade NOVA and Universidade Aberta who were motivated to collaborate with us. During this collaboration, according to the collaborative DMP-building method, the data steward verified that they had no experience either in DMP creation or in RDM in general. Also, they had no contact with the DPO, so the data steward helped researchers in various aspects during the creation of the DMP, such as repository choice, collaboration with other departments (IT staff of their institutions), and ethical and property issues (owner and data publication). No issues were raised with respect to the costs. During the preliminary analysis of this project, we assumed that it was classified by the FOS as "computer and information sciences, history and archaeology", which was fully confirmed. Also, we did not make a mistake in assuming the type of data in the project. The project contains only data that can be published without any restrictions. As a result of the four meetings and several email exchanges during seven months, we published the first version of the DMP [147] and we discussed the need to monitor the plan during the development of the project. During this collaboration, another project **"LifeSkillsVR"** was proposed by the PI, which also required the creation of a plan. As this issue was quite important and urgent for the PI and for the project in general, the data steward took it up for consideration. However, this project is not part of our sample and is considered and analyzed separately from the sample projects. Chapter 8, Section 8.2 describes the similar projects that are not part of the sample but were taken by the data steward as part of their regular work related to the creation, monitoring or improvement of DMP.

FuturePharm - Pharmaceutical supply chain of the future

The project led by the University of Coimbra received funding from the FCT in 2021 to develop optimization models for pharmaceutical supply chains. Following

4 https://frontowns.fcsh.unl.pt/

the collaborative DMP-building method and first scenario, the data steward sent an email to the PI. However, they did not receive a reply. After several unsuccessful attempts to communicate with the PI, contact was established through the joint efforts of the data steward and a university professor, who had collaborated with the PI before on other projects. The PI indicated the most suitable researcher for this collaboration, and the data steward went ahead with interviews and the creation of the first version of the DMP. As in most cases, the researchers had neither experience in creating plans, nor were they in contact with the DPO when preparing the grant proposal. During the four months of the collaboration, a first draft of the DMP was created, where the data steward helped with repository choice and organization of the data management on the project in general. Cost issues, additional documents, collaboration with other departments, and ethical and property issues were not raised. Our preliminary assignment about the scientific domain was partially correct. We assumed that this project is classified under the FOS as "computer and information sciences, economics and business; other medical science" domain. However, it was classified as "other engineering sciences and technologies". This inaccurate preliminary assignment is due to the project's link to pharmaceutical production. It led the data steward to assume that there would be much more private and confidential data in the project, and additional agreements would be needed to make them public. This assignment was partially correct and helped to discuss with researchers more specific information about the embargo period. Specifically, although the data did not contain any private and confidential information, the researchers were not prepared to release their data without an embargo period. Currently, the DMP is under revision, waiting for the last corrections until its publication.

ATTEST - Advanced Tools Towards cost-efficient decarbonisation of future reliable Energy SysTems

According to the first scenario, the data steward sent an email to the PI of the project approved in 2020 and financed by Horizon 2020[5] to understand their situation with respect to the DMP, knowing that Horizon 2020 requires DMP submission. The principal goal of this project is to develop and operationalize a toolbox to support transmission system operators and distribution system operators of energy networks[6]. The PI confirmed that DMP creation is one of the deliverables of the project, and as they have created it already, they are more interested in the data deposit and description support than in DMP creation. However, during the collaboration, the data steward proposed an analysis of their DMP and verified that the plan looks more like a Project Management Plan, and the project actually has a lot of sensitive and confidential information, and therefore the plan is not published and without monitoring. In this context, the data steward proposed the creation of the DMP-light version for easier monitoring, updating and keeping as a "live" document with the possibility of publishing it on Zenodo. The data steward also helped to improve the description of the tools used for data collection, the expected size of the data deposit and publishing, to identify other types of data different from sensor data such as agreements and reports, which were not seen as data by project members. Moreover, during the five months of the collaboration, the data steward verified that the PI had contact with the DPO of the INESC TEC to clarify issues related to the data policies. Our preliminary assignment the scientific domain was correct. The project is classified as the "electrical engineering, electronic engineering, information engineering; other engineering and technologies" domain[7]. The data steward did not assume that the project collected sensitive data. However, the telemetric data were collected by real energy networks and included data from cus-

5 https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/programmes/h2020
6 https://www.inesctec.pt/pt/projetos/attest
7 https://cordis.europa.eu/project/id/864298

tomers. The DMP-light version was very useful for this project and is being under revision by all project participants. The PI is motivated to publish it and sees it as the output of a project that could be shared with the community.

EUniversal - Market enabling interface to unlock flexible solutions for cost-effective management of smarter distribution grids

This project is funded by Horizon 2020 in 2020 and is related to the development of a new multi-energy and multi-consumer concept guaranteeing sustainable, secure and stable electricity supply[8]. To help with DMP creation, the data steward contacted the PI by email, following the first scenario. During the collaboration with the PI, the data steward verified that the PI had no contact with the DPO. However, the first version of the DMP had already been created and submitted on the funders' platform. The PI was interested in analyzing the current version of the DMP, whereby the data steward could point out aspects for improvement. The data steward suggested improvements related to the repository and storage issues, organization of datasets and defining more specific details in the DMP, such as licenses, and restrictions related to data. Moreover, the data steward emphasized that the plan is still in the early stages of development, requiring detailed information. This was not the case for the DMP-light. Since INESC TEC is not the project coordinator and some partners have support in filling their part of the DMP from DMP support services, it became necessary to clarify the type of support that the PI wanted from the data steward. It can be: 1) support in creating DMP for the whole project; 2) support in describing DMP only for INESC TEC part; or 3) cooperation between support services of the other project members for joint creation of DMP. This verification could be added to our DMP-building method and seen as one of the improvements of the method when implementing the DMP support system. After one month of collaboration, one meeting and several emails, the DMP is in revision, awaiting PI's response. Our preliminary assignment related to the domain was correct; the project is classified as "electrical engineering, electronic engineering, information engineering; other engineering and technologies" domain according to FOS[9]. It was not very clear to the data steward whether personal data was being collected in pre-analysis and needed more exploration along with researchers.

TRUST-AI - Transparent, Reliable and Unbiased Smart Tool for AI

A project related to the development of a transparent, reliable, and unbiased tool focusing on machine learning applications has been approved for funding by Horizon 2020 in 2020[10] and led by INESC TEC. Collaboration between the data steward and the PI started at the beginning of the project and followed the collaborative DMP-building method, with the data steward sending an email to the PI offering support in DMP creation. This case is one of ten case studies where we applied our collaborative method for training and is described in Chapter 5. Since then, the first DMP has already been created and submitted on the funding platform, and many questions related to the repository, ethical issues, identification of collected data, and responsibility have been clarified. The researchers did not have much experience with RDM issues. Before working with the data steward, they had created an ethical agreement with the support of the DPO at INESC TEC, so the data steward used this document as a basis for developing the first draft of DMP. At present, after nineteen months of collaboration, three meetings and a series of email exchanges, the planned monitoring session has taken place in June 2022. As the data steward has been accompanying the project practically from the beginning, the monitoring session went quite easily, adding more details and specific information about the

8 https://www.inesctec.pt/pt/projetos/euniversal
9 https://cordis.europa.eu/project/id/864334
10 http://www.trustai.eu/

data, their management and changes related to them to the already existing plan. The gaps that were left with the note "to be described in more detail in the next version of the DMP" were also filled in. This point shows us that stating as much information as possible in the first version of the plan with a note to improve it in the next version helps to reduce the time for analysis and the monitoring session itself. It also helps to avoid forgetting important aspects to be specified in the plan while improving its quality. Moreover, since project members did not want to publish their current version of the plan, the data steward suggested creating a light version of the DMP for publication, which interested the PI. The DMP-light is now under review by the PI in order to improve and apply the changes that occurred in the project during this time. Our preliminary assignment about the scientific field has turned out to be partly correct: the project is classified as "computer and information sciences and political science" according to the FOS classification, rather than "computer and information sciences and health"[11]. This did not prevent the data steward from making correct preliminary assignments about the type of data in the project, and from being more attentive to the analysis of the existence of public, sensitive, personal and private data in the project.

NOVATERRA - Integrated novel strategies for reducing the use and impact of pesticides, towards sustainable mediterranean vineyards and olive groves

The NOVATERRA project started in 2020 with funding from the Horizon 2020 and aims to reduce the use of controversial plant protection products (also known as pesticides)[12]. Following the collaborative method, after contacting the PI, the data steward found that a DMP had already been created, with the help of the data manager of one of the project partners. The data steward offered to analyze the created plan for possible improvement and prepare a light version of the DMP for publication, as their existing DMP was being disseminated exclusively at a confidential level without publication. Two months of collaboration, one meeting and an exchange of emails resulted in suggestions for improvement related to the repository and in the proposal for the creation of the DMP-light version. Moreover, researchers were advised to change the Google Drive that researchers used for data storage, as it was not very suitable for storing confidential data. The preliminary assignment that the project refers to "computer and information sciences; other engineering and technologies; other agricultural sciences" is not quite correct, as it is defined as "agriculture, forestry and fisheries; biological sciences" according to the FOS classification[13]. Therefore, the data steward made a slightly wrong assignment about the type of data collected and created by the project, concluding that the project would only collect public data, while the researchers planned to collect and create public data, data with some restrictions and private data with some confidential business-related information. However, during the collaboration with the PI, all specific aspects related to these types of data were clarified. Overall, the analysis of their DMP showed a good level of detail, responsibility and motivation of all project partners in this task, recognizing the value and importance of RDM issues. The DMP is now submitted to the funders' platform and awaits the next monitoring session in line with the deliverable schedule. The data steward made sure that researchers could contact them at any time to get support and is awaiting a decision from the PI to create a DMP-light version.

11 https://cordis.europa.eu/project/id/952060
12 https://www.inesctec.pt/pt/projetos/novaterra
13 https://cordis.europa.eu/project/id/101000554

PRySM - PRecision Sprayer Ground Robot

This project started in 2020 with funding from Horizon 2020 and aimed to develop a modular and precise spraying robot for spraying tasks in mountain vineyards[14]. Email contact with the PI enabled a meeting to be made, during which the data steward found out that the project had already been finalized, the funder had not requested a DMP, and therefore there was no contact with either the DPO or the data steward to create a plan. Although the first email was sent by the data steward while the project was still in progress, the project was already at the completion stage, and as the DMP was not mandatory, the researchers did not need support in creating the DMP, so the first email went unanswered. Our preliminary assignment about the FOS classification was partially correct, the project belongs to the "agriculture, forestry, and fisheries; social and economic geography; electrical engineering, electronic engineering, information engineering"; the data steward indicated that the project belongs to "computer and information sciences; other engineering and technologies; other agricultural sciences" domain. However, we could not verify our preliminary assignment about the type of data, as the data steward did not proceed with the analysis of the project. Although the PI has experience of creating DMPs and data management in general, they are interested in further collaboration for new projects, now knowing where to request support.

SCORPION - Cost effective robots for smart precision spraying

The Scorpion project aims to improve spraying efficiency and develop a modular unmanned tractor (robotic platform)[15]. It started in 2021, is funded by Horizon 2020 and is led by INESC TEC. They already have a first version of the DMP, where the DPO helped clarify issues related to ethical and data protection issues. However, their DMP is only distributed to partners without publication. The data steward analyzed their version of the plan and made several suggestions for improvement regarding internal data sharing and access, defining responsibilities and adding more detail on datasets. In addition, despite the fact that researchers insisted they would not collect personal data, the analysis of the context of the project allowed the data steward to confirm that personal data would be created during the video recordings of experiences. In this regard, the data steward recommended the creation of informed consent for personal data. Developing the DMP-light version of their plan was also proposed for easier monitoring, updating and publication on Zenodo. Although the PI already had experience in RDM and on creating DMPs, they continued to have difficulty with managing the data during and after the project and saw this collaboration as very supportive. They were also motivated to get the DMP-light version, recognizing the benefits of publishing, and sharing their plan with the community. As a result of six months of collaboration, two meetings and several email exchanges, the DMP-light version was created and sent to the PI for revision. According to the FOS classification, this project refers to "agriculture, forestry, and fisheries; Social and economic geography; electrical engineering, electronic engineering, information engineering" and the data steward preliminary assignment was partially correct "computer and information sciences; other engineering and technologies; other agricultural sciences"[16]. Thus, in our view, this led the data steward to a partially correct assignment about the data collected on the project. Instead of "public, sensitive, personal (because of the videos), restricted" the data steward assumed "public, private or restricted".

14 https://www.inesctec.pt/en/projects/prysm
15 https://scorpion-h2020.eu/
16 https://cordis.europa.eu/project/id/101004085

CircThread - Building the Digital Thread for Circular Economy Product, Resource & Service Management

The project has been approved for Horizon 2020 funding in 2021, and is aimed at the creation of the software platform for an "information broker" in the market to unlock access to product data between stakeholders who are currently in silos, and use it to improve Circular Economy decision-making throughout the product lifecycle[17]. Following the first scenario and the collaborative DMP-building method, the data steward sent an email to the PI to offer support in DMP creation. During the meeting, it was agreed that the PI would check whether any created version of the DMP existed in the project. It was also agreed that the data steward would start creating a first draft of the plan, which would be useful in any case. If the DMP does not exist, then the PI will send our created plan for evaluation by project members; if it does, then our plan will help improve the existing one or be used as a DMP-light version for publication. After the verification of the plan's existence, the PI sent to the data steward the existing DMP. Researchers created the plan with the support of the DPO by one of the project members and did not want to publish it. However, during the four months of collaboration, the data steward already had a draft DMP, so the existing DMP was analyzed, some comments for improvement were suggested, and the draft DMP created by the data steward was proposed as a light version of the DMP recommended for publication. The PI is very interested in collaborating with the data steward, as he has no experience in RDM issues, and feels it is important to have a person who can help on the project. He invited the data steward to be present in the project meetings, because he thinks that only the data steward can explain all the ideas and clarify issues. He also invited the data steward to participate in other projects that are in preparation. Our preliminary assignment related to the scientific domain was partially correct, instead of "environmental engineering; economics and business" the data steward supposed "computer and information sciences; economics and business"[18]. That is why the data steward assumed that the project will collect more private and sensitive data, not personal. The DMP-light and comments on the improvement of the main DMP are in the process of revision. In general, the data steward has helped with issues related to the definition of the type of data collected/created during the project, storage, repository, metadata, additional documents related to sensitive and personal data and data organization in general.

EIT_RIS_Hubs_2021 - EIT Manufacturing RIS Hubs 2021

This project is linked to the creation of a contact point for EIT Manufacturing in the EIT RIS countries to support local manufacturing innovators, started in 2021 and funded through Horizon 2020[19]. The data steward sent several emails to the PI, only to find out some time later that the PI had been replaced by another person. However, contact with the new PI also required connection with another researcher at INESC TEC. Following the collaborative DMP-building method, the data steward did arrange a meeting with the new PI, who did not have any experience in RDM and who initially saw no interest in creating a plan, as they thought they would not be collecting any data. The data steward explained that the creation of this DMP could be seen both as training for new projects and as a more proper and detailed organization of the internal data within the project. Using the specific situation of a change of the PI as an example, the data steward showed that the new PI coming into the project neither knew how data management in the project

17 https://circthread.com/
18 https://cordis.europa.eu/project/id/958448
19 https://www.inesctec.pt/pt/projetos/eit_ris_hubs_2021

was organized, nor the responsibilities, and if there was a plan, it would have been much easier to sort out the data and other RDM issues. During the collaboration, the data steward helped with data organization and storage issues. Moreover, when the PI mentioned that different online training sessions and courses would be developed during the project, the data steward pointed to the possibility of sensitive and personal data collection. Thus, our preliminary assignment regarding data turned out to be partially correct. The data steward did not initially assume that personal data could be collected, as the data steward did not know about the courses. The presence of personal data in the project was particularly emphasized in the first version of the plan, which was created during the three months of cooperation. The project belongs to the "economics and business; computer and information sciences" according to FOS classification; our preliminary assignment was correct. At the moment, the DMP is in the revision process by the PI. During the collaboration, the PI also suggested other projects, such as **"TURING"**, **"IESMA Summer School"**, **"IMMC"**, **"Green APS"**, **"GreenMA"**, **"ConFacts"**, **"Tech2Market"**, and **"Demo4Green"**, where researchers might be interested in receiving support from the data steward in DMP creation. As these projects are not part of our sample, they are analyzed separately and described in Chapter 8, Section 8.2.

Inno4Vac - A cloud-based systems-immunology platform for reliable predictions of vaccine efficacy

A project involving the use of artificial intelligence combined with big data and computational modelling to create an open-access, cloud-based platform for in silico vaccine efficacy evaluation and development has been approved for funding in 2021 by funder Horizon 2020[20]. This project has several sub-projects, one of which INESC TEC is responsible for. The data steward contacted the PI to offer assistance with DMP creation and discovered that the plan already exists, but needs monitoring and improvement. Furthermore, the DMP is defined as a confidential document without publication and each project partner is responsible only for their own data activities and fills in only their allocated part of the plan. Therefore, the PI, although experienced in RDM, was interested in having the data steward review the plan and suggest improvements if any were required. During the collaboration, the data steward verified that the DMP had already been submitted on the funders' platform, and the funder requested more detail on this document, due to the project collecting some pharmacy data, which are sensitive, private, and require specific attention on their management. Researchers also had contact with the DPO for clarification of issues and review of the consortium agreement. Although the project is related to health, our preliminary assignment related to the scientific domain was partially correct; the project is classified as "computer and information sciences; health sciences; basic medicine; mathematics" according to FOS, instead of the "medical biotechnology, computer and information sciences; biological sciences; health sciences"[21]. However, as in general, these domains typically collect sensitive, personal, and private data, the data stewards' assignment about the data was totally correct. As a result of 6 months of collaboration, the data steward proposed some recommendations to improve the plan related to the data storage and access issues, licenses, additional documents and collaboration with the DPO that sensitive and personal data almost always require. The data steward sent all recommendations to the PI and is waiting for a response not only on monitoring and improvements but also on the possibility of creating a DMP-light version for publication.

20 https://www.inesctec.pt/en/projects/inno4vac
21 https://cordis.europa.eu/project/id/101007799

PAFSE - Partnerships for science education

Project PAFSE is a research and education project aimed at solving public health problems, which started in 2021 and is funded by Horizon 2020[22]. Following the first scenario, the data steward sent an email to the PI and offered support in DMP creation. Several emails remained unanswered, so the data steward, in cooperation with another collaborator, tried to contact the project PI in other ways. After the first meeting, the PI confirmed that he was very motivated and interested in our collaboration, but the number of emails received in a day made it difficult to filter out important proposals. Although INESC TEC is not the main coordinator of this project and is not responsible for the results related to the creation of the DMP, and the DMP has already been created, the data steward offered their assistance in analyzing the existing version of the plan and comments on its improvement, if necessary. In addition, during the plan review, the data steward realized that the researchers had already established contact with the DPO and had a researcher in charge of DMP and RDM issues in the project, who in turn supported the creation of the DMP. The data steward proposed some improvements related to data organization, storage issues, and informed consent. Researchers were advised to change the Google survey forms, which are not the best for personal data, and verify some issues with the DPO. Analysis showed that their DMP was well-detailed but not published, so the data steward suggested publishing the DMP and meeting with the researcher in charge of DMP creation. Our preliminary assignment that the project is classified as "educational sciences; computer and information sciences; health sciences" according to the FOS is partly correct, the project belongs to the field of "political sciences; mathematics; earth and related environmental sciences; health sciences"[23]. However, this did not affect the identification of the type of data; during the project, researchers collected all types of data: public, sensitive, personal, and private. Currently, after one meeting, several email exchanges and three months of collaboration, the first version of the DMP was created and submitted to the funders' platform, comments for improvement are being finalized and the data steward is awaiting PI's decision to publish the DMP. Finally, during this collaboration, the PI also applied for support for another project "**VR2CARE**", which is not on our sample list. Because of its urgency and importance, the data steward accepted to support it; this project is described and analyzed in Chapter 8, Section 8.2.

FIRE_RES - Innovative Technologies and Socio-Ecological-Economic Solutions for FIRE RESilient Territories in Europe

The FIRE-RES project funded by Horizon 2020 started in 2021 and is developing a holistic and integrated fire management strategy to efficiently and effectively address Extreme Wildfire Events in Europe in 11 Living Labs thanks to its Innovation Actions[24]. During the preparation of the project proposal, one of the researchers (not the PI) sent an email to the data steward asking for support in the preparation of the basic DMP that was requested by the funding organization together with the application (second scenario). Moreover, researchers also had support from the DPO at INESC TEC regarding data and consortium agreement. After the project received funding, the data steward sent several emails to clarify the point of the development of the DMP. The emails remained unanswered, so the data steward, through another collaborator, contacted the PI to clarify the need for a DMP and further collaboration. The PI of the project confirmed that there is a data manager who supports them in the DMP creation. However, they were not opposed to the data steward analyzing the existing plan and giving their comments on improvements. Thus, the data steward collaborated with the data manager and suggested

22 https://www.inesctec.pt/pt/projetos/pafse
23 https://cordis.europa.eu/project/id/101006468
24 https://fire-res.eu/

Table 20: Summary of the 14 projects that have been supported by the data steward in DMP creation

DMP title	Scientific domain	Budget (K euros)	DMP status
Qualtos: Quality Assurance in Long Term Observation Systems	Other engineering ant technologies	≈ 100	In revision
FronTowns: Think big on small frontier towns: Alto Alentejo and Alta Extremadura leonesa (13th - 16th centuries)	Computer and information sciences; history and archaeology	≈ 190	Published
FuturePharm: Pharmaceutical supply chain of the future	Other engineering sciences and technologies	≈ 250	In revision
ATTEST: Advanced Tools Towards cost-efficient decarbonisation of future reliable Energy SysTems	Electrical engineering, electronic engineering, information engineering; other engineering and technologies	≈ 4000	In revision
EUniversal: Market enabling interface to unlock flexible solutions for cost-effective management of smarter distribution grids	Electrical engineering, electronic engineering, information engineering; other engineering and technologies	≈ 8000	In revision
TRUST-AI: Transparent, Reliable and Unbiased Smart Tool for AI	Computer and information sciences; political sciences	≈ 4000	In revision
NOVATERRA: Integrated novel strategies for reducing the use and impact of pesticides, towards sustainable mediterranean vineyards and olive groves	Agriculture, forestry, and fisheries; biological sciences	≈ 5000	Submitted on funders' platform
PRySM: PRecision Sprayer Ground Robot	Agriculture, forestry, and fisheries; social and economic geography; electrical engineering, electronic engineering, information engineering;	≈ 190	Project closed without DMP
SCORPION: Cost effective robots for smart precision spraying	Agriculture, forestry, and fisheries; social and economic geography; electrical engineering, electronic engineering, information engineering;	≈ 2300	In revision
CIrcThread: Building the Digital Thread for Circular Economy Product, Resource & Service Management	Environmental engineering; economics and business	≈ 8000	In revision
EIT_RIS_Hubs_2021: EIT Manufacturing RIS Hubs 2021	Economics and business; computer and information sciences	≈ 605	In revision
Inno4Vac: A cloud-based systems-immunology platform for reliable predictions of vaccine efficacy	Computer and information sciences; health sciences; basic medicine; mathematics	≈ 2000	In revision
PAFSE: Partnerships for science education	Political sciences; mathematics; earth and related environmental sciences; health sciences	≈ 1425	In revision
FIRE_RES: Innovative Technologies and Socio-Ecological-Economic Solutions for FIRE RESilient Territories in Europe	Sociology; economics and business; political sciences; earth and related environmental sciences	≈ 21500	In preparation

some improvements related to the sensitive data that would be collected. Without any meetings, only by exchanging emails, during the 17 months from the proposal preparation, the DMP is under preparation, where the data manager is responsible for the creation of the first detailed version of the DMP, and the data steward is available for additional support at any moment. Our assignment related to the scientific domains is partially correct, where the data steward indicated "environmental engineering; earth and related environmental sciences; other social sciences", instead of the "sociology; economics and business; political sciences; earth and related environmental sciences" domain according to the FOS classification[25]. The data stewards' assignment related to the data was correct. Researchers collect all types of data: public, sensitive, personal, and private.

We have summarized these projects in Table 20, with the name of the project, its scientific domain, budget, and DMP status.

7.5.2 Projects considered in stand-by with respect to the DMP

The projects we describe next are those which have been contacted by the data steward according to the first scenario, but either the PI has not shown interest in creating a DMP or is still in the process of analyzing the support proposal.

THOR - Computer Assisted Thoracic Assessment using POCUS[26]

During the pre-analysis of this case, the data steward verified that the PI had some support from the DPO at INESC TEC in clarification of issues during the proposal development phase, a preliminary assignment that the PI will probably accept our support. However, all data steward attempts to contact the PI were unsuccessful, but by continuing to try to meet with the PI, the data steward made some other important contacts that led them to one of the researchers on this project, with whom they held a meeting and explained the support proposal for DMP creation. The researcher showed interest in collaboration and decided to pass this information to the PI. The PI can contact the data steward, in case of interest, but the data steward will also periodically remind the researcher about this topic which is not closed yet. Thus, the project is considered in stand-by with respect to the DMP.

CAGED - Computer Assisted Gastric Cancer Diagnosis[27]

This project is also associated to the PI of the THOR project. In this case, the data steward also found that the DPO provided support to clarify issues at the proposal

25 https://cordis.europa.eu/project/id/101037419
26 https://www.inesctec.pt/en/projects/thor
27 https://www.inesctec.pt/en/projects/caged

creation stage, and contacted one of the researchers of this project. The meeting has been held, the researcher has given all the information to the PI and the project is considered in stand-by with respect to the DMP, waiting for contact from the PI.

Connect2Oceans - Connecting Atlantic and Arctic Oceans to Decipher Climate Change Impact on Plankton Microbiome Functions[28], ATLANTIS - The Atlantic Testing Platform for Maritime Robotics: New Frontiers for Inspection and Maintenance of Offshore Energy Infrastructures[29] and FIRELOGUE - Cross-sector dialogue for Wildfire Risk Management[30]

All attempts to establish contact with the PIs of these projects for 5-7 months were unsuccessful. By taking different ways to establish collaboration, the data steward was able to meet with the PI of another project who had some connection with these projects. But during the meeting, it was pointed out that these projects hardly need support for DMP creation. However, the information was passed to the projects' PIs. At present, the project is considered in stand-by with respect to the DMP, awaiting some contact from the PIs.

MATinMOL - Matter Waves in Moiré Lattices[31]

Contact with the PI of this project was established through third-party contacts. Although the PI was interested in cooperation, it was revealed at the meeting that INESC TEC was not the main coordinator of the project, so the decision to collaborate and support DMP creation would be taken by another institution. However, the final answer was that they do not need to create any DMP for this project, but are interested in further cooperation and will contact the data steward when needed.

BeFresh - On incorporating consumer behaviour into the supply chain planning of fresh products[32]

Various attempts to contact the PI of this project by e-mail, as well as through third parties, have led to the answer that they are not interested in creating a plan for this project, as it is not mandatory.

MYTAG - Universal environmental monitoring devices based on Intelligent optical scattering analysis amplified with molecularly imprinted polymers[33] and inSITE - Insitu ore grading system using LIBS in harsh environments[34]

The PI of these two projects left all of the data stewards' attempts to establish collaboration unattended. Despite this, the data steward was able to hold a meeting with several of the researchers of these projects. During the meetings, they showed interest in creating DMPs, planning to pass all the information about this opportunity of support to the PI of the project in order to get authorization to start the collaboration. Moreover, they requested the creation of the DMP for another project **"Submerse"**, which is described in Chapter 8, Section 8.2. At the moment, the data steward is waiting for a decision on these two projects, so the project is considered in stand-by with respect to the DMP.

28 https://www.inesctec.pt/en/projects/connect2oceans
29 https://www.inesctec.pt/en/projects/atlantis
30 https://www.inesctec.pt/en/projects/firelogue
31 https://www.inesctec.pt/en/projects/matinmol
32 https://www.inesctec.pt/en/projects/befresh
33 https://www.inesctec.pt/en/projects/mytag
34 https://www.inesctec.pt/en/projects/insite

UNEXUP - UNEXMIN Upscaling[35]

After several unsuccessful attempts to contact the PI of this project, the meeting was arranged with the help of another researcher. During the meeting, the PI confirmed that they were interested in such collaboration, but since the plan is not mandatory for this project and they cannot allocate the time required for its creation, they will refuse to collaborate for the moment and keep in contact for the next project. Moreover, during the meeting, they suggested another project, ***"DeepField"***, that could benefit from our support. The DeepField project is out of our sample, so it is analyzed and described in Chapter 8, Section 8.2.

AI_REGIO - Regions and DIHs alliance for AI-driven digital transformation of European Manufacturing SMEs[36]

The PI of this project replied almost immediately that they were not interested in creating a plan and did not need our support at the moment. However, in preparation for future projects they will contact the data steward. Pre-analysis showed that this project had support from the DPO at INESC TEC related to the creation of the partner agreements.

OneNet - One Network for Europe[37]

Contact with the PI of this project was established only through contact with another researcher, all emails from the data steward were left unanswered. During the meeting the PI was very interested and motivated to cooperate and enthusiastically talked about the project. However, the collaboration did not take place. INESC TEC is not the principal coordinator of this project, there is a partner in charge of the DMP creation. In general, the data steward offered two options for DMP support. The data steward can analyze the existing DMP and offer comments to improve it, or create a DMP-light version suitable for publication, since the existing DMP version is not planned to be public. The PI needed to request authorization to collaborate and decide what type of support is more interesting in this case. However, project members decided not to use our support. The PI will keep in contact with the data steward for future projects.

HumanE-AI-Net - HumanE AI Network[38]

The PI responded to the email and indicated a researcher from the project with whom the data steward scheduled a meeting. During the meeting, the data steward realized that the researcher was not interested in creating a DMP, but needed support in depositing and describing datasets. So, the data steward contacted the data manager at INESC TEC, who is currently responsible for the INESC TEC research data repository, and then scheduled another meeting directly with the PI to better understand the need for a DMP for the project. During the second meeting, the PI was very interested in collaboration, but continued to focus on data deposit because a DMP was not required for this project. Finally, it was decided that the PI would contact the data steward when they started preparing the new project.

35 https://www.inesctec.pt/en/projects/unexup
36 https://www.inesctec.pt/en/projects/ai_regio
37 https://www.inesctec.pt/en/projects/onenet
38 https://www.inesctec.pt/en/projects/humane-ai-net

Table 21: Summary of the 15 projects where the data steward established contact but DMP support did not advanced

Project title	Scientific domain	Budget (K euros)
THOR: Computer Assisted Thoracic Assessment using POCUS	Computer and information sciences; health sciences	≈ 240
CAGED: Computer Assisted Gastric Cancer Diagnosis	Computer and information sciences; medical biotechnology	≈ 250
Connect2Oceans: Connecting Atlantic and Arctic Oceans to Decipher Climate Change Impact on Plankton Microbiome Functions	Computer and information sciences; biological sciences; other natural sciences; environmental engineering, environmental biotechnology	≈ 250
ATLANTIS: The Atlantic Testing Platform for Maritime Robotics: New Frontiers for Inspection and Maintenance of Offshore Energy Infrastructures	Environmental engineering; biological sciences; electrical engineering, electronic engineering, information engineering	≈ 7050
FIRELOGUE: Cross-sector dialogue for Wildfire Risk Management	Sociology	≈ 3260
MATinMOL: Matter Waves in Moiré Lattices	Physical sciences	≈ 245
BeFresh: On incorporating consumer behaviour into the supply chain planning of fresh products	Computer and information sciences; economics and business	≈ 250
MYTAG: Universal environmental monitoring devices based on Intelligent optical scattering analysis amplified with molecularly imprinted polymers	Computer and information sciences; other natural science	≈ 250
inSITE: Insitu ore grading system using LIBS in harsh environments	Computer and information sciences; physical science	≈ 1940
UNEXUP : UNEXMIN Upscaling	Environmental engineering; other engineering and technologies	≈ 3000
AI_REGIO: Regions and DIHs alliance for AI-driven digital transformation of European Manufacturing SMEs	Computer and infromation sciences; economics and business; political sciences	≈ 9200
OneNet: One Network for Europe	Electrical engineering, electronic engineering, information engineering; other engineering and technologies	≈ 22000
HumanE-AI-Net: HumanE AI Network	Computer and information sciences; political sciences	≈ 12000
EUSCORES: EUropean - Scalable and Complementary Offshore Renewable Energy Sources	Environmental engineering	≈ 34850
MAGPIE: sMArt Green Ports as Integrated Efficient multimodal hubs	Environmental engineering; sociology; other engineering and technologies; social and economic geography	≈ 25000

EUSCORES - EUropean - Scalable and Complementary Offshore Renewable Energy Sources[39]

The PI responded quickly, saying that they were not interested in support for DMP creation, and would contact the data steward by email or in person when the need arises.

MAGPIE - sMArt Green Ports as Integrated Efficient multimodal hubs[40]

The PI of this project responded quickly and did not want our support because they already had support from the data manager. Also, they were not interested in having the data steward review the existing plan and comment on improvements.

A summary of these projects can be seen in Table 21 with the project name, research area, and budget.

7.5.3 Projects without an established contact

Next, we will describe projects from which the data steward could not get any answer, not only by emails, but also through other ways of contact.

PassCert - Exploring the Impact of Formal Verification on the Adoption of Password Security Software[41]

Over the course of seven months, the data steward sent several emails and made various attempts to meet with the PI. However, all of the data stewards' attempts to establish contact with the PI were unsuccessful.

BetterFactory - Grow your manufacturing business[42] and MARI4_YARD - User-centric solutions for a flexible and modular manufacturing in small and medium-sized shipyards[43]

During the six months, the data steward made different tentative to contact the PI of these projects by email, through the other researchers and collaborators, without

39 https://www.inesctec.pt/en/projects/euscores
40 https://www.inesctec.pt/en/projects/magpie
41 https://www.inesctec.pt/en/projects/passcert
42 https://www.inesctec.pt/en/projects/betterfactory
43 https://www.mari4yard.eu/mari4_yard-project-launch/

Table 22: Summary of the 4 projects where the data steward could not get any response to by all possible ways

Project title	Scientific domain	Budget (K euros)
PassCert: Exploring the Impact of Formal Verification on the Adoption of Password Security Software	Computer and information sciences	≈ 70
BetterFactory: Grow your manufacturing business'	Economics and business; Computer and information sciences	≈ 8000
MARI4_YARD: User-centric solutions for a flexible and modular manufacturing in small and medium-sized shipyards	Sociology; economics and business; electrical engineering, electronic engineering, information engineering; mechanical engineering	≈ 5900
RISC2: A network for supporting the coordination of High-Performance Computing research between Europe and Latin America	Computer and information sciences; other engineering and technologies; economics and business	≈ 514

Table 23: Summary of the 6 projects where the data steward proposed creation of the DMP-light version

Project title	Scientific domain	Budget (K euros)	DMP light
ATTEST: Advanced Tools Towards cost-efficient decarbonisation of future reliable Energy SysTems	Electrical engineering, electronic engineering, information engineering; other engineering and technologies	≈ 4000	In revision
OneNet: One Network for Europe	Electrical engineering, electronic engineering, information engineering; other engineering and technologies	≈ 22000	Proposed
TRUST-AI: Transparent, Reliable and Unbiased Smart Tool for AI	Computer and information sciences; political sciences	≈ 4000	Proposed
NOVATERRA: Integrated novel strategies for reducing the use and impact of pesticides, towards sustainable mediterranean vineyards and olive groves	Agriculture, forestry, and fisheries; biological sciences	≈ 5000	Proposed
SCORPION: Cost effective robots for smart precision spraying	Agriculture, forestry, and fisheries; social and economic geography; electrical engineering, electronic engineering, information engineering;	≈ 2300	In revision
CIrcThread: Building the Digital Thread for Circular Economy Product, Resource & Service Management	Environmental engineering; economics and business	≈ 8000	Proposed

success.

RISC2 - A network for supporting the coordination of High-Performance Computing research between Europe and Latin America[44]

As a result of five months of contact the PI by the data steward, as well as contact through other collaborators and researchers, the collaboration never happened. All attempts were unsuccessful.

A summary of these projects can be seen in Table 22 with the project name, research area, and budget.

Finally, we created a Table 23 with projects where DMP-light version was proposed or created.

The result for all projects is presented in Figure 87, which indicates the total number of projects in the sample **(33)**, the number of projects with which we were able to establish collaboration and support them in the creation of plans **(14)**, the number of projects with which we were able to contact, but for some reason the plans were not created **(15)**, the number of projects with which we were unable to establish contact **(4)**, and the number of the projects where DMP-light version were proposed or created **(6)**. After completing the collaboration, we sent out a survey[45] that help us analyze the results of our data steward support, understand the researchers' difficulties, their needs and, based on this, suggest improvements.

7.6 SUMMARY

In order to propose a DMP support system in an institution, it is necessary to apply and systematize the collaborative DMP-building method in a larger number of projects from different scientific domains. For this purpose, in this chapter, we define our sample of **70** projects and apply different filters such as "Year", "Funder" and data of the end of the project, resulting in **33** projects. We also carry out a pre-analysis of these projects in order to gather the necessary information in advance to familiarise ourselves with their domains, and to understand what type of data could be collected. Preliminary analysis helps the data steward to focus on specific points during collaboration with researchers, where there may be difficulties, or the

44 https://www.inesctec.pt/en/projects/risc2
45 https://tinyurl.com/mw6y2vz6

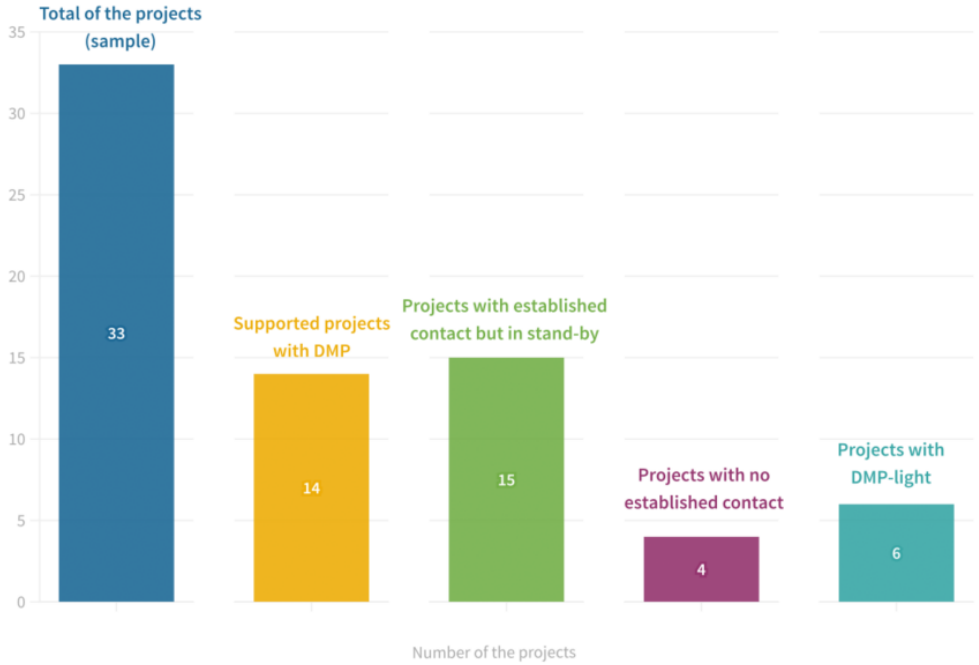

Figure 87: Summary of the projects analyzed in our sample

need for help from other stakeholders or institutional departments, or the creation of additional documents. In this chapter, we also propose two different scenarios for establishing collaboration between the data steward and researchers, which can be used as a guide for any data steward in any institution to support researchers. Although we have only one project that requests support according to the second scenario, we expect that when the DMP support system is implemented there will be more such projects. Thus, both scenarios should be taken into account in the institution when developing a DMP support system. This chapter also defines the metrics for evaluating DMP support. Finally, a description of all projects in our sample, the process of support by the data steward, and data related to collaboration with projects are also presented here. This description also shows the effort of finding ways to collaborate with projects that required persistence, activity, and systematicity on the part of the data steward.

In the next chapter, we discuss the results obtained during the systematization of our collaborative DMP-building method. The collected data are analyzed, and data steward support and the collaborative method are evaluated according to the established metrics, survey answers and researchers' feedback. Moreover, additional projects from previous work and new ones suggested are also analyzed.

8 ANALYSIS AND EVALUATION OF THE COLLABORATIVE METHOD

In order to evaluate the application of the collaborative DMP-building method for its systematization and the effectiveness of the support based on it, as well as to suggest improvements and the final version of a well-designed institutional DMP support system, in this chapter we analyze the results obtained from various sources. More specifically, we analyze the results according to the established metrics in Chapter 7, Section 7.4. We also analyze the results obtained from the answers to the survey[1] that was sent to researchers after our collaboration, the informal researchers' feedback obtained during the collaboration and email exchanges, as well as the data stewards' general view of the difficulties faced, ways to solve them and other aspects related to the DMP support system. This analysis and its results help us determine the efficiency of our developed collaborative method, which is the basis for the institutional DMP support system, its benefits, importance, and the importance of having a data steward. We analyze if our collaborative method could be used in any project with any budget, for any scientific domain, and if the DMP support system that is being developed and proposed could be used in any institution.

Moreover, in this chapter, we show that during the collaboration with the researchers from sample projects selected for analysis, new proposals sometimes emerge. They are not included in the sample analysis, but they are additional information that, in one way or another, affects the development and proposal of the DMP support system. Together with the new projects, we also present here the projects described in Chapter 5 which are under monitoring. These additional projects also help to summarize the results of the sample analysis.

8.1 ANALYSIS OF THE RESULTS

First of all, the analysis of the data we collected during the work, presented on the dataset "Assessment of metrics for the development of an institutional DMP support system" [131] and illustrated in Figure 87 indicates that the data steward was able to establish contact with **14** projects out of **33** of the projects in the entire sample and supported them in the DMP creation. The Figure also shows that we were able to establish contact with **15** projects, but their plans were not created ("stand-by"). Moreover, the Figure presents that with **4** projects, the data steward was not able to establish any contact; and for **6** projects, a DMP-light version was proposed or created. Thus, we can state that only **14** projects out of **33** in the entire sample expressed interest in receiving support for the DMP creation. Since this is less than half of the projects, in the course of the analysis, we look at the reasons for the rejection or "stand-by" of the collaboration in order to identify possible difficulties and overcome them in the future. However, we highlight that during the collaboration - we received several requests for DMP support for other projects not included in our sample, which are described in Section 8.2. They could help broaden the analysis of results by showing that interest may increase when researchers become aware of existing DMP support or are already collaborating with the data steward to improve their plans.

1 https://tinyurl.com/mw6y2vz6

8.1.1 Analysis of quantitative metrics

To evaluate the ***Quant1 Number of acceptance (scenario 1)*** and ***Quant2 (Number of requests (scenario 2))*** metrics, we analyzed data in our dataset [131], which was filled during the collaboration with researchers. We counted projects that accepted collaboration with the data steward.

The first metric - ***Quant1 (Number of acceptance (scenario 1))*** indicates the total number of projects in the sample that have accepted the DMP-related support offered by the data steward following scenario 1 where the data steward contacted the researchers. This metric also helps to understand the researchers' interest in supporting the creation of a DMP, calculated as the percentage of projects that have accepted DMP support from all projects in the sample. The value of this metric is **13** projects out of **33** total of the projects in our sample.

The second metric - ***Quant2 (Number of requests (scenario 2))*** indicates the total number of projects which required DMP-related support from the data steward. In this case, we have only **1** project that requested DMP support following scenario 2. Although this value is very small, it is expected because most of the projects in our sample were contacted by the data steward. But at the same time, we also supposed that the sample may include projects that have already requested support from the data steward before and fall under the second collaboration scenario. This metric is important to analyze in Section 8.2, when we describe projects which were suggested by PIs during the collaboration or existing ones from previous work which are under monitoring, to see whether interest grows when researchers are aware of the institutional support for DMPs.

These initial results indicate that currently, researchers are more interested in DMP creation support when the data steward contacts them. However, as was mentioned above, this could be explained by a lack of awareness of the existence of this type of support within the institution. It was observed that the more the data steward communicates with the researchers, the more interested they are in getting the DMP support, using the second scenario, contacting the data steward when necessary. However, this interest extended more to further projects than to the ones in our sample. In some cases, this was due to the fact that the creation of a DMP for the project was optional and not required by the funders, and also due to time constraints. In other cases, this was due to the fact that a plan had already been created by one of the project partners. In addition, several projects were at the final stage and the creation of the DMP was not mandatory. Thus, these initial results show that a well-established DMP support system in an institution should use both scenarios, which will help to organize the collaboration between the data steward and researchers. The results show that the more often funders require mandatory DMPs, the more researchers will look for DMP support. Moreover, the researchers will look for this support at their institutions, which should implement DMP support systems and RDM services in general to meet the needs of both researchers and funders. Establishing and implementing a DMP support system in an institution where researchers can formally request support for the DMP creation will greatly help them meet existing RDM requirements. Thus, the growing need to create plans increases the demand and interest for support for their creation.

We describe together the results of the analysis of the metrics ***Quant3 (Sensitive, personal, private data issues)***, ***Quant4 (Repository issues)***, ***Quant5 (Costs issues)***, ***Quant6 (Additional Documents)***, ***Quant7 (Collaboration with other stakeholders)*** and ***Quant8 (Ethical and property rights issues)*** because all of them show how many projects have had difficulties with these issues, and how many projects the data steward has helped to solve them. To evaluate these metrics we analyzed data in our dataset [131], which was filled during the collaboration with researchers. In the following, we provide values for each of these metrics.

The ***Quant3 (Sensitive, personal, private data issues)*** metric shows that **6** projects out of **14** projects that received DMP support in our sample encountered difficul-

ties related to this type of data and resolved them with the data steward support. For example, in some cases, the researchers did not know at the beginning that they would collect a personal type of data and did not know that DPIA would be required. In other cases, the researchers thought that only public data would be collected, but during the creation of the DMP, it appeared that these public data could only be opened after the embargo had expired. The metric shows that almost half of the projects faced difficulties or issues related to this type of data. In fact, training the collaborative DMP-building method on the ten case studies described in Chapter 5 also showed that this topic is challenging for researchers. This type of data almost always requires additional documentation or collaboration with other stakeholders, and sometimes the understanding that these data are being collected may only come, for example, in the middle of a project. The systematization of the collaborative method only emphasises this situation and the importance of support from the data steward during the project lifecycle.

The ***Quant4 (Repository issues)*** metric shows that **13** projects out of **14** that received DMP support in our sample, obtained help in repository-related issues and on choosing the more appropriate one for each project, as well as clarification of issues related to depositing and publishing both data and DMPs. Moreover, the results show that many researchers are not aware of existing repositories, neither of their certification status and trustworthiness. However, collaboration with the data steward has broadened their knowledge, and has made them aware of the institutional repository and the support provided by INESC TEC in dataset organization, description, and citation. Thus, **7** projects out of **14** that received DMP support in our sample chose the INESC TEC research data repository to deposit their datasets. In many cases, the INESC TEC RDM repository has been specified in the DMP, as each dataset deposited there is assigned a DOI. These results show that the selection of the repository is a very important aspect and that the data steward can help researchers by selecting the most suitable one from the beginning of the project.

The ***Quant5 (Costs issues)*** metric shows that **1** project out of **14** that received DMP support in our sample have had difficulties with cost issues such as purchasing disks for data preservation, storage and other technology-related costs. Although, during the collaboration, the data steward consulted many researchers on some budgetary issues related to RDM and mentioned the importance of these calculations in advance, in our case only **1** project needed such help. From our point of view, this is due to the already prepared and approved project budget. In other words, we assume that this support would be requested more likely during the proposal preparation stage than after the project is approved by the funder. Thus, the data steward support is important at the initial planning stage to avoid unplanned costs, for example.

The ***Quant6 (Additional Documents)*** metric illustrates that in **6** projects out of **14** that received DMP support in our sample, the data steward identified the need for additional project documents, such as DPIA, informed consent, and agreements between project members or for the Ethics Committee. This, in turn, helps to anticipate the difficulties related to collecting sensitive, personal or private data during the project and avoid unforeseen situations. Thus, for example, in one project, the data steward identified the need to create the informed consent because during data collection they recorded various videos in which workers could be unintentionally filmed. In another example, due to changes in the data sharing rules established at the beginning of the project, it became necessary to create a DPIA and analyze the risk associated with personal data. This result is similar to the result of the Quant3 (Sensitive, personal, private data issues) metric because they are linked. As was mentioned above, this type of data almost always require additional documents.

The ***Quant7 (Collaboration with other stakeholders)*** is also connected to the previous metrics because additional documents and collection of sensitive, personal, or private data most often require collaboration with other stakeholders such as the DPO. Thus, in our case, Quant7 reveals that in **8** projects out of **14** that received

DMP support in our sample, the data steward helped researchers make contact with the DPO team, which is also useful for subsequent projects. Moreover, the analysis of these aspects helps us to understand how we can improve the internal collaboration processes between the DPO team and other stakeholders, internal regulations related to project management in the institution, and in the future, suggest a method that will help to facilitate collaboration with the Ethics Committee. During the collaboration, the data steward also emphasized the importance of paying more attention to projects that collect sensitive, personal and private data. As practice shows, the need for data steward support can arise at any stage of the project. We propose that when the data steward creates the first draft of the DMP, leave a note stating "*During the project in accordance with the data collection, their processing and analysis the need to create the DPIA and collaboration with the DPO may emerge. More detailed information about this aspect will be added in the following versions of the DMP*" in the plan. This could help researchers to be more attentive to this topic and remember to analyze this aspect when monitoring the plan.

The ***Quant8 (Ethical and property rights issues)*** metric shows that **9** projects out of **14** that received DMP support in our sample have received help from the data steward related to the ethical, legal, copyright, Intellectual Property Rights, owner and license issues. In the course of collaboration with researchers, it was observed that the researchers considered this issue important and tried to pay enough attention to it beforehand. This is why some of the projects in our sample requested the assistance of the DPO team already when creating the proposal (**8** projects out of **33** requested support from the DPO before our collaboration). The data steward also raised these issues during meetings with the researchers in order to avoid difficult situations in the future and to discuss these issues with all project members as early as possible before data collection and publication. Thus, one of the researchers confirmed: "*Yes, most of the time we don't think about these issues beforehand. But, in fact, it is important to discuss it with all project members in advance. It is good that you remind us of this*".

The last quantitative metric to analyze is related to the DMP creation time - ***Quant9 (Time)***. This metric shows us that the DMP creation process, its monitoring and other improvement sessions require a lot of time and effort. The values for this metric are available in our dataset [131] and were filled during the collaboration with researchers. For each project, we recorded the time the data steward needed for collaboration for the DMP creation. The collaboration with researchers has confirmed that the creation of a DMP is not a continuous process. All tasks related to the DMP occur at various intervals. For example, there are intervals to discuss some technical issues with all project partners, or some intervals to organize meetings with partners to check the plan, analyze it, improve it, and approve it for publication. The results of the data analysis show that the time required to create the plans, monitor, improve and keep them as a *"living"* document varies between **1** and **19** months of two-way collaboration work with some intervals as mentioned above. This result also shows that the DMP is created, monitored and updated throughout the project and sometimes after its conclusion. In most cases, it takes **1-7** months to create a DMP, whereas monitoring and improvement take **7-19** months. These time intervals show the need and importance to create the DMP at the very beginning of the project. Because creating the plan at the end of the project by doing it in a hurry can not only compromise the quality of the plan itself but also be unhelpful to the project as a whole. Publication of the plan, however, most often depends on the dimension of the project, its budget and the number of partners in the project. According to the results, the more complex the project, the more time it takes for DMP creation and publication.

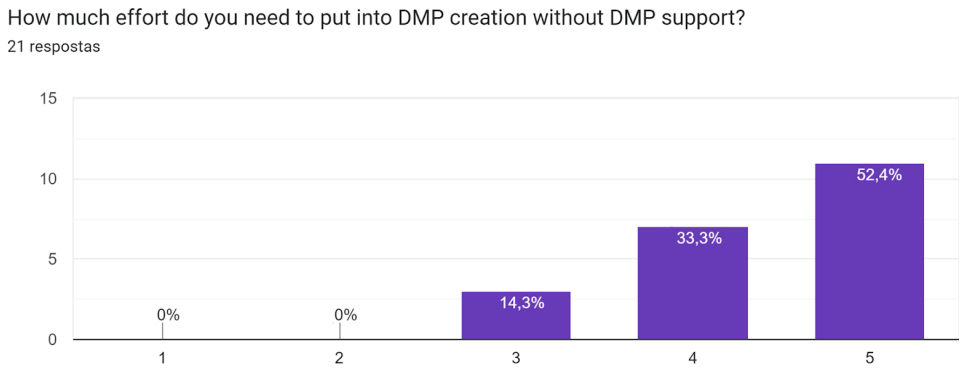

Figure 88: Qual2 - Effort that researchers need to put into DMP creation without DMP support

8.1.2 Analysis of qualitative metrics

In addition to quantitative metrics, qualitative metrics ***Qual1 (Motivation)***, ***Qual2 (Effort)***, ***Qual3 (Researchers Satisfaction Score)***, ***Qual4 (Rejection)***, and ***Qual5 (Impact)*** were also identified for analysis.

The first metric ***Qual1 (Motivation)*** defines a general ideia of the researchers' motivation and their readiness to perform DMP-related tasks. To assess this metric, we analyze feedback from researchers that we received during meetings and collaborations. Namely, during the course of the collaboration, the data steward tried to analyze how researchers interact, their motivation, interests, activity in meetings, and performance of DMP-related tasks. If researchers were motivated to collaboration we marked this project with "Yes" on our dataset [131] in the Qual1 (Motivation) column, otherwise "No". As a result, we can confirm that the majority of researchers (**12** projects out of **14** that received DMP support) were motivated to collaborate and create/monitor/improve their plans. Some of them were very grateful, and invited the data steward to accompany them on new projects. They also added the data steward as a member of the current project team, confirming that such work should be officially recognized and appreciated. Others have requested data steward support for further projects and those of their colleagues who were probably not yet aware of the existing RDM institutional support. Moreover, the support from the data steward was included in one of the budgets of the new project proposal being prepared.

The work also shows that researchers are more motivated to create a DMP during the first two meetings, expecting it to be a quick process, and then motivation diminishes. It is, therefore, very important to try to make the DMP creation process as easy and quick as possible. Our collaborative method, where the data steward creates the first draft of the plan, is exactly what they expect, reducing the time researchers spent gathering the information for the plan and also creating its structure.

The ***Qual2 (Effort)*** shows how much effort researchers need to put into DMP creation without data steward support and their perception of the weight of this task without any help. The metric is based on the researchers' responses to the two survey questions *" How much effort do you need to put into DMP creation without DMP support?"* and *"How long do you thing it would take you to create a DMP without Data Steward support (hours, days, weeks, etc.)?"*. For the first question was used a scale of **1** to **5**, where **1** is little effort, **5** is a huge effort (Figure 88). To answer the second question, the researchers had to provide their options for answering the question.

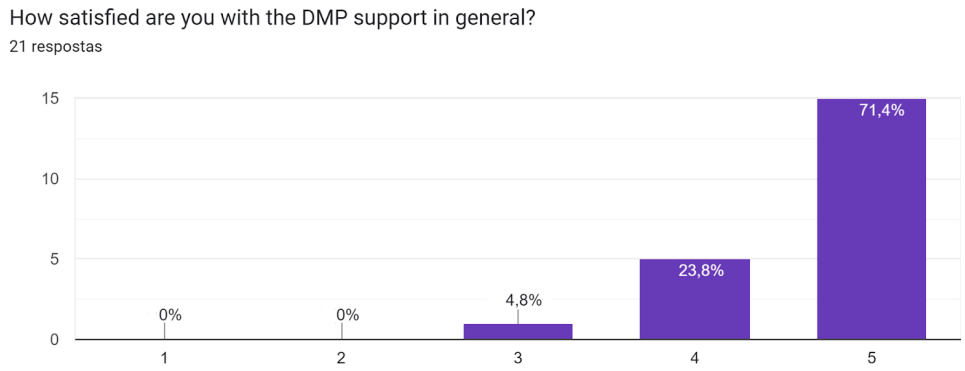

Figure 89: Qual3 - Researchers Satisfaction Score

The results show that the majority of survey respondents (52,4%) indicated that without support, DMP creation would require a huge effort. As for the time needed to create it, the researchers indicated varying responses, where the interval ranged from a few days to **12** months, or they might be unable to estimate the time. Some researchers declared that they could create a DMP in a month, but this would require improvements; others stated that they spread the effort over several project partners, so it is quite difficult to say how much time each partner would need. Overall, the results show that the more effort it takes for researchers without any support to create a DMP, the more important the DMP support services are.

The next metric - ***Qual3 (Researchers Satisfaction Score)*** helps us to understand how satisfied researchers are with DMP support and how easy it is to create a DMP with the data steward support. The data for analyzing this metric were obtained from researchers' answers to the survey question: *"How satisfied are you with the DMP support in general?"*. Figure 89 shows that the satisfaction level of most researchers was **"Completely satisfied"**, which on the graph corresponds to the value **5**. The value **1**, in this case, means **"Absolutely unsatisfied"**. As the results show, no one chose the answers **"1-Absolutely unsatisfied"** and **"2 - Unsatisfied"**; one respondent chose the answer **"3 - Indifferent"**; and five respondents - the answer **"Satisfied"**. The majority of the respondents were **"Completely satisfied"** with DMP support.

The following ***Qual4 (Rejection)*** metric shows that **19** projects out of **33** projects in the entire sample rejected our collaboration, did not respond to the data steward or kept the collaboration in "stand-by". The results for this metric we obtained in our dataset [131], which was filled during the collaboration with researchers, counting the projects with rejection. We defined this metric to analyze the reasons for rejection, because it can help us to improve the DMP support system, internal processes and methods related to institutional support. This analysis is based on the data steward perception obtained during the conversation with researchers and their informal feedback.

In general, we received different reasons for rejection, but we could highlight the following:

1. The DMP is not mandatory for the project, and the PI does not have the time or capacity to deal with this issue personally at the moment;
2. The DMP is not mandatory for the project, and none of the researchers of the project has time or opportunity to deal with the issue at the moment;
3. The DMP is not mandatory for the project, the project is coming to an end, and the PI does not feel the need to create it;
4. The DMP is mandatory, has already been created, and there is no interest in our support;

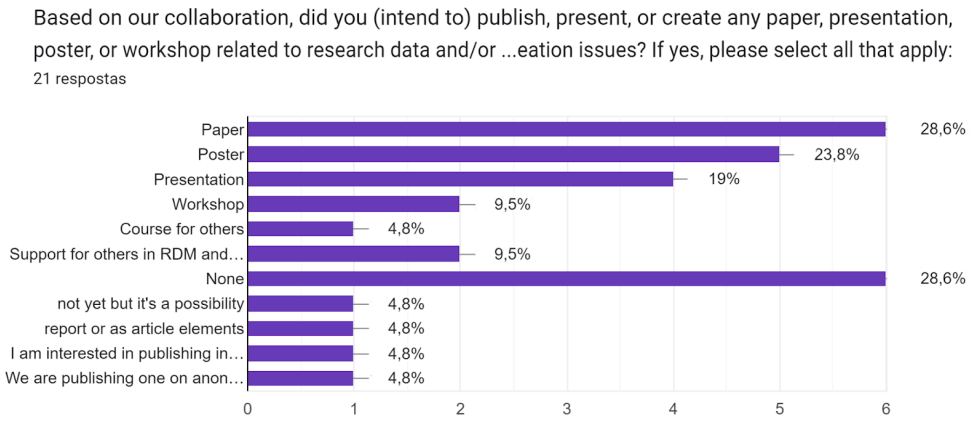

Figure 90: Qual5 (Qualitative impact). Intention to disseminate the DMP

5. The DMP is mandatory, but the responsibility for creating the plan lies with another project member, and there is no interest in our support;
6. Other unknown motives.

Summarizing the results, the perception of the data steward is that most of the rejections were due to a lack of time and motivation. Also, it could be due to the fact that the DMP creation is not mandatory by funders and also that an institutional DMP support system with an approved policy regarding the plans is not yet officially implemented. These results may change if plans become mandatory for submission.

The last qualitative metric for analysis is ***Qual5 (Qualitative impact)***. It is based on the researchers' answers in the survey and shows how they benefit from our collaboration beyond the created DMP and practice in its creation. The first question *"Based on our collaboration, did you (intend to) publish, present, or create any paper, presentation, poster, or workshop related to research data and/or DMP creation issues? If yes, please select all that apply:"* checks whether the researchers are interested and motivated in disseminating and promoting the newly obtained knowledge related to RDM and DMP issues. In other words, we analyze the researchers' intention to publish and create papers, presentations, workshops, and other similar activities related to the DMP creation issues. As can be seen in Figure 90, we have only **6** responses where the researcher has not published and does not intend to publish, present or create any DMP-related material and **1** response where the researcher has not published anything yet but is planning to do so. The other answers show that most researchers are interested in sharing their knowledge and experiences related to DMP creation and RDM in general. Moreover, in this question, the researchers could write their answer option, so in addition to the proposed answers, they also indicated *"report or as article elements"*, *"I am interested in publishing in the various types described. But in a co-author position..."*, and *"We are publishing one on anonymization procedures, not DMP"* (see full answers in Appendix G). By analyzing these responses, we can say that interest in DMP and RDM issues is growing regardless of the scientific domain, the number of publications related to these topics is increasing, and researchers are more engaged in RDM issues. They have started to share their experiences, which allows us to analyze their needs and the challenges they face during the execution of the RDM tasks, thus helping to improve our tools and methods.

The second question that researchers responded to is *"What did the data steward do for your project concretely?"*. In this question, they could choose different options for the answer and also propose their own (Figure 91). As the results show, the majority of respondents indicated that the data steward helped to "*make DMP of*

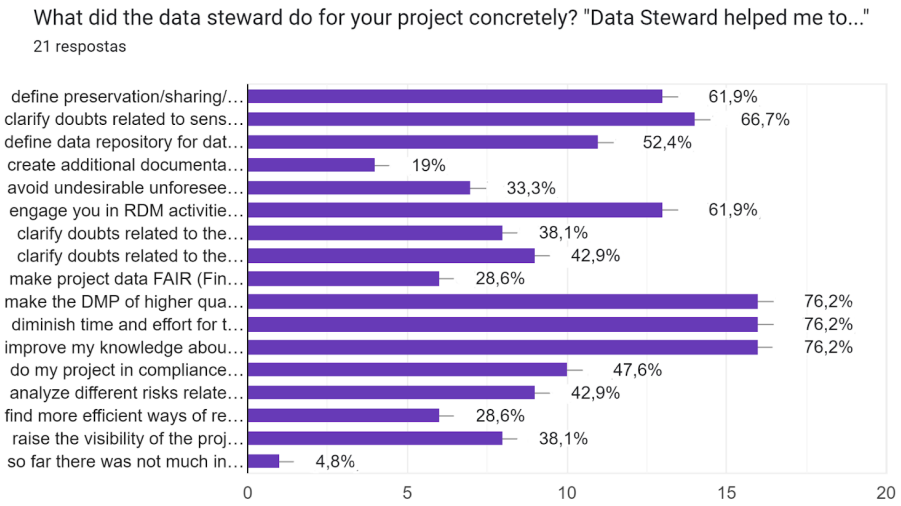

Figure 91: Qual5 (Qualitative impact). The type of data steward support

higher quality and more detailed" (**76,2**%), "*diminish time and effort for the DMP creation"* (**76,2**%) and "*improve my knowledge about managing my data of the project"* (**76,2**%).

One of the responses was as follows: "*so far there was not much interaction given that this is a recent service. Hopefully, this will change in the next versions and projects*". This answer can be explained by the fact that, indeed, the DMP support service has not yet been officially implemented, and is in development. Therefore, this answer is fairly obvious.

There were also other answers where the researchers generally confirmed that the data steward helped to clarify various questions related to sensitive, personal, private data, and other RDM issues, to define RDM strategies and rules during the project as well as to avoid undesirable unforeseen events and to predict possible problems in the initial stages of the project (see full answers in Appendix G).

Furthermore, during the collaboration, the data steward analyzed this Qual5 metric according to the three profiles defined and described in Chapter 5. More specifically, the first profile is a regular researcher, who can be a PI or someone in charge of the RDM issues in a project. The second profile is a researcher who acts as the data steward in a research group. The third profile is a collaborator of the institution who, in one way or another, intersects with research projects and, therefore, with RDM issues. This could be a project manager, funding officer, DPO, or others [145].

The result of the analysis shows that for the ***first profile***, the qualitative impact of DMP creation support is focused on reducing the effort and time required from the researchers, creating well-detailed DMP and keeping it as a*"living"* document, improving researchers knowledge of RDM and DMP issues. They are not interested in becoming trainers or organizing workshops but in improving their knowledge, getting DMP support for their benefits and creating DMPs easier and faster.

Second profile researchers are more interested in detailed information about the DMP-related processes and involvement in RDM issues, and see this collaboration as training which can help them in the future to create workshops and courses related to RDM in their scientific domain. Moreover, they are interested in presenting themselves as data steward for their institution or project groups and providing RDM and DMP support when needed. Confirmation of this is seen in answers presented in Figure 90, where some researchers indicated that they intend to use this collaboration in workshops, courses for others and support to others in RDM issues. Some of them use the collaborative DMP-building method as a basis which

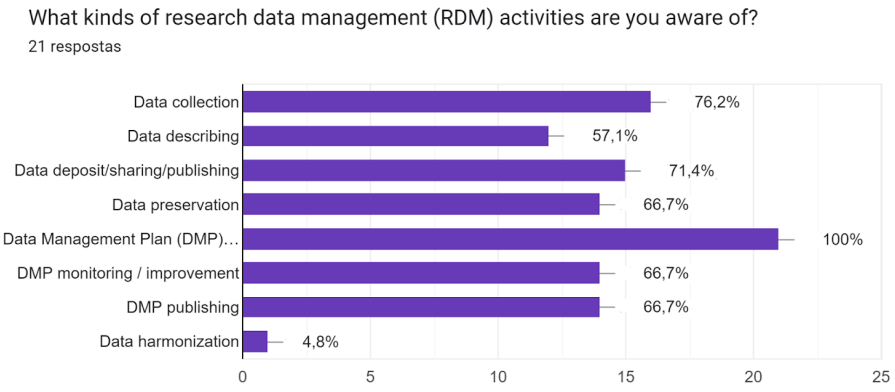

Figure 92: Awareness of RDM activities among researchers

verifies that it is reusable and suitable for any scientific domain and any institution for DMP-related processes.

The analysis of results according to the ***third profile*** shows that collaboration of the data steward with other stakeholders is also very important during the DMP support. Moreover, all institutional stakeholders and departments benefit from this collaboration, facilitating and complementing each other's work and helping to resolve difficult and unforeseen situations related to RDM; thereby improving the quality of the project management process, showing better results for the institution as a whole and enhancing its competence and compliance with existing RDM requirements from funders and the scientific community at large. In other words, the data steward can be seen as a central element bringing together all the stakeholders that are involved in one way or another in RDM.

8.1.3 Analysis of researchers' feedback

To get broader results regarding researchers' awareness and experience with RDM issues, we also included a variety of RDM-related questions in our survey and collected the researchers' feedback. This survey was sent to all projects participants with whom the data steward collaborated, so several project members (not just PIs) could respond in one project. The survey resulted in 21 responses.

Although the survey answers show that most researchers are aware of different RDM activities (Figure 92), analysis performed by the data steward during the collaboration and first meetings showed that only 7 researchers have experience in RDM tasks [131]. In our opinion, this is caused by the distance between theory and actual practice. The topics of RDM are already discussed in various sources. However, there is still a lack of application in practice. This conclusion is also confirmed by the answers presented in Figure 93, where - **71.4**% of the researchers indicated that they had no experience with DMP creation. Those who had experience in creating plans gave the following answers: *"I tried to fill out a DMP document, but I found it very difficult; I had a collaboration for the creation of a DMP"*, *"I had previous experience but not with so much detail and rigor"*, *"I created a DMP supported by the institutional data steward"*, *"...supported by the use of DMP tools, like DMPOnline"*, *"...using the information from the workshops/webinars/masterclasses on the creation of DMP that I attended"*, and *"...analyzing different samples and existing plans"*. This leads us to state that data steward support is needed not only for DMP creation but in RDM issues in general.

Moreover, the results show that prior to our collaboration, **66.7**% of researchers were unaware of any institutional support (Figure 94), but most of them indicated the importance and value of data steward support in DMP creation (Figure 95). The

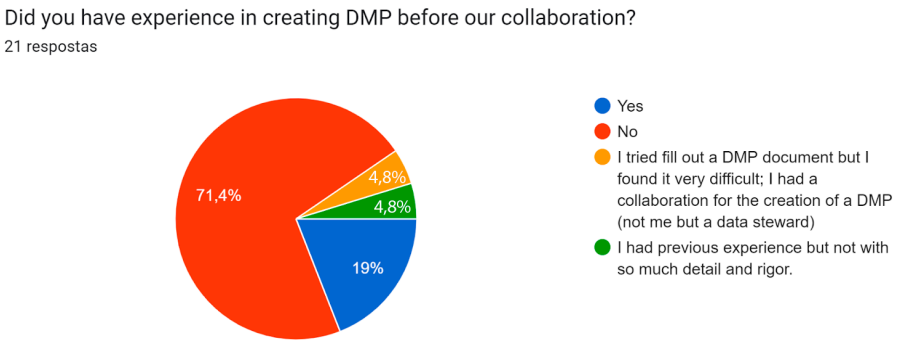

Figure 93: Researchers' experience in DMP creation

researchers' answers to the question: *"In your opinion, how valuable and important was the support given by the data steward during the DMP creation "* helped to deepen the analysis of their opinion about the DMP support satisfaction in general. The results show that they consider the DMP support received from the data steward as **"Very important and valuable"**, which corresponds to a value of **5** in Figure 95.

Also, all respondents unanimously mentioned the need for institutional support for all RDM activities (Figure 96). Some researchers emphasize the importance of the existence of a data steward, providing feedback: *"We are so busy with our project that we forget the tasks we have to do concerning DMP and RDM in the project. It is good that the data steward keeps an eye on this and gets in touch with us often, remembering the situations to resolve"*.

During the course of the work, it was observed that the more the data steward met and communicated with the researchers, explaining the goals of collaboration and the development of the DMP support system for the institution, the more researchers became interested in obtaining support for future projects. Some researchers commented: *"We will soon be creating a new proposal for funding, so, we would very much like to count on your support both in creating the proposal and in the DMP that will be presented to the funder after the project is approved"*.

In some cases, researchers stated that as they did not know that the institution has support for DMP creation and that the DMP was not mandatory for their project, they were trying to "*escape*" from this task because they had difficulties answering all the questions. Also, there are situations where researchers report that they do not want to create a DMP and believe that DMP support must always be provided when needed. They compare this support with the functioning of the IT department and with services like Help-Desk or Service-Desk when users receive support for any matter related to IT issues.

Furthermore, some researchers confirm that the existence of the DPO is a positive factor. However, they recognize that the DPO team cannot help in such detail in the creation of DMP as a data steward using the proposed collaborative DMP-building method. Hence, they are very motivated by the development of the institutional DMP support system that manages to guarantee support both in project proposals and in their follow-up after funding approval.

The **100**% importance of the existence of institutional support for RDM activities (Figure 96), and not the **100**%, but **81**% importance of the data steward presence in the institution (Figure 97) can be explained by the lack of knowledge of the special terminology by the researchers. Not all researchers have named a data steward as a "data steward"; in some cases, the data steward was called as the person in charge of RDM support at INESC TEC. In our work, we use the term "data steward" to refer to individuals who support researchers in DMP creation and other RDM issues. However, there is another term that can be used in this case - "data curator". In

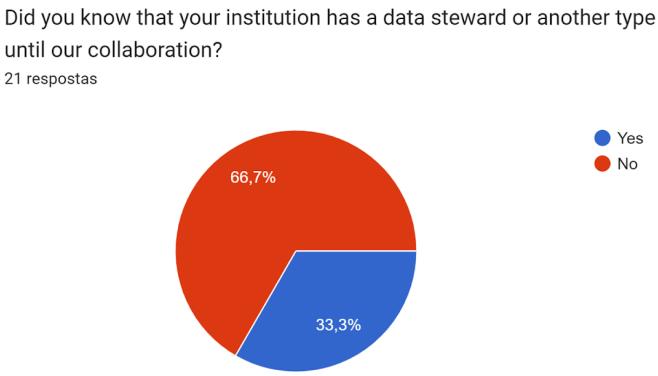

Figure 94: Awareness of any institutional support for RDM

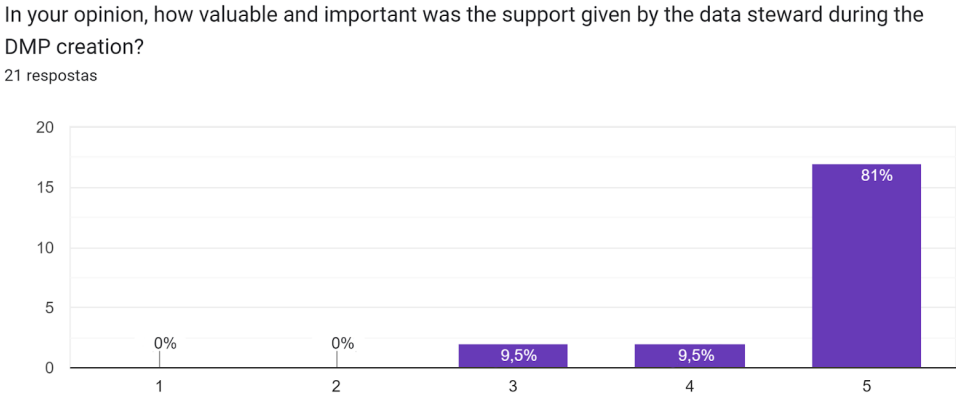

Figure 95: The importance and value of the data steward support in DMP creation

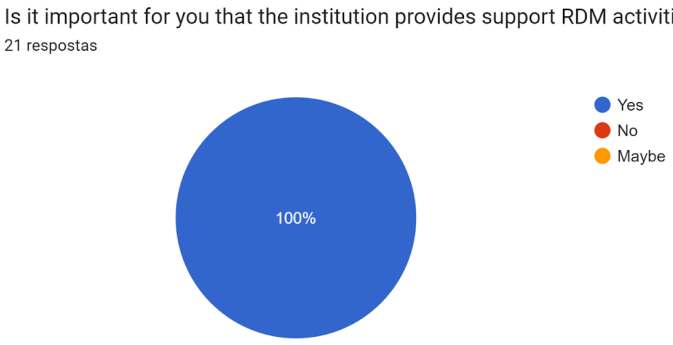

Figure 96: The importance of the institutional RDM support

our view, both the data curator and the data steward are involved in the organization of the RDM infrastructure and can support researchers in various RDM tasks. However, the data curator is more involved in data-related processes such as data description, and the data steward is more involved in collaboration with researchers and other institutional stakeholders. In our case, the data steward performs all the functions of both data curator and data steward, supporting researchers in all RDM issues. This, in turn, indicates the need for a formal definition of the data steward

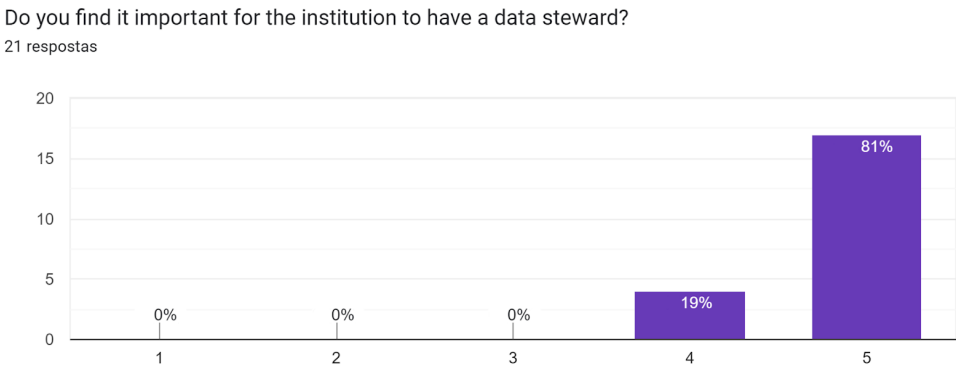

Figure 97: The importance of the existence of the data steward in the institution

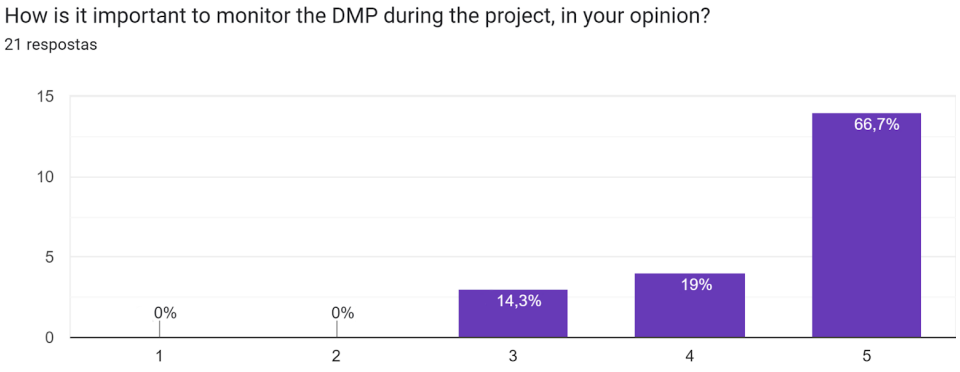

Figure 98: Importance of the DMP monitoring

position at the institution and approval of the DMP support system and related processes in institutional workflows.

Part of the questions in the survey aimed at analyzing the need for subsequent monitoring of the DMP. As DMP monitoring in projects usually takes place within **4-6** months after its creation, most of the monitoring in our sample will be outside the scope of this work. As such, we wanted to get the researchers' views in advance to understand their needs, difficulties, and requirements and propose a more adequate method for the DMP monitoring process as a part of the DMP support system under development. Figure 98 shows that **66,7**% of respondents indicated DMP monitoring as **"Very important"**, which corresponds to the value **5** of our question. At the same time, **85,7**% of researchers indicated that this process requires support (Figure 99) and that DMP should be monitored every **6** months.

The question about the time interval for DMP monitoring was an "open" question that allowed researchers to provide their own answers. So there are different responses were received, such as: *"once a year"*, *"3-4 times during the project"*, *"depend on the project dynamics"*, or *"Ideally continuously. In practice, depends on project dynamics. Usually, it happens once or twice a year at most"*. This variation in responses shows that the researchers understand the necessity of monitoring the plan and reflecting on any changes that occur during the course of the project. Also, their professional experience shows that it is better to monitor changes **3-4** times during the project, or more often depending on the dynamics of the project, which coincides with the already existing recommendation in the scientific community.

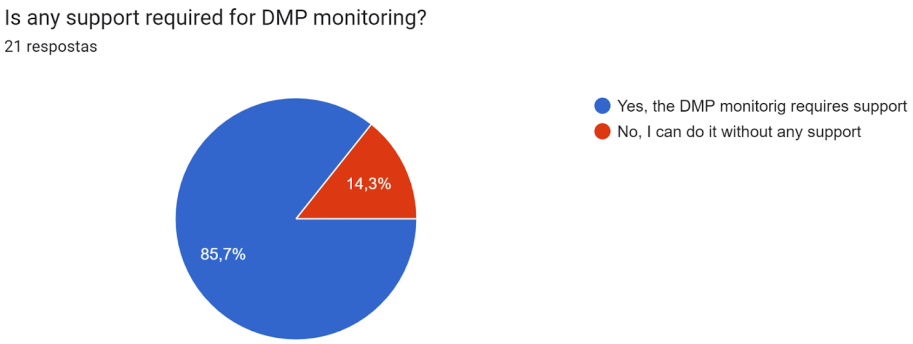

Figure 99: Importance of the DMP monitoring support

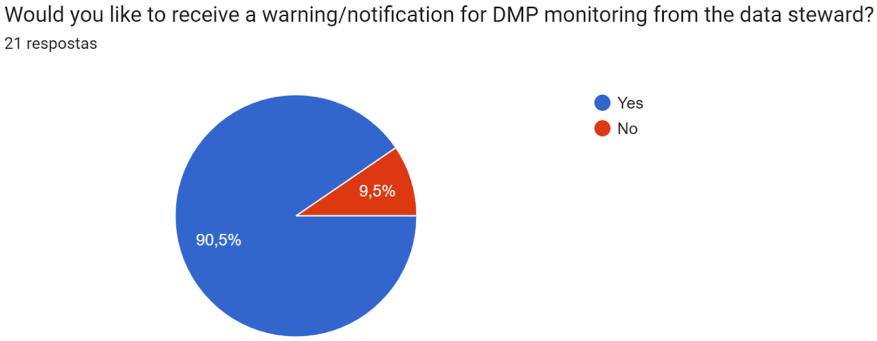

Figure 100: Researchers' answers related to the notifications for DMP monitoring

We also asked researchers if they wanted to receive any notifications from the data steward for the DMP monitoring process. **90.5**% of researchers indicated **"Yes"**, confirming the importance of having a data steward at the institution who can follow each project through to completion and remind researchers to keep their plan as a*"living"* document (Figure 100).

Regarding flexible metadata models, controlled vocabularies, and data domain protocols that are designed to assist in the DMP creation process, the researchers noted during the DMP creation process that they believe some scientific domains have specific needs. Those differences may be useful for one domain, but not meaningful for another. For example, they highlighted education, health, artificial intelligence and historical domain, and others where sensitive or personal data is most commonly collected. As such domains require more careful attention to additional documents such as DPIA, analysis of specific rules of sharing, e.g. when data from people who have been dead for over 500 years are collected, or ethical issues related to artificial intelligence. Some of the researchers stated that there are differences, but they find it difficult to identify them. The answers that there are no differences are also present.

This can be explained by the fact that the identification and creation of the controlled vocabularies and proposal of the Domain Data Protocols is a difficult and time-consuming process. We described this process in detail in Chapter 5, Section 5.4, where controlled vocabularies for the Psychology domain were created and tested and the Domain Data Protocol was proposed. Also, in the course of collaboration with researchers from other domains, we have identified several possible terms that can later be tested as controlled vocabularies and then used to propose a Domain Data Protocol, that, in turn, will help in DMP creation. More specifi-

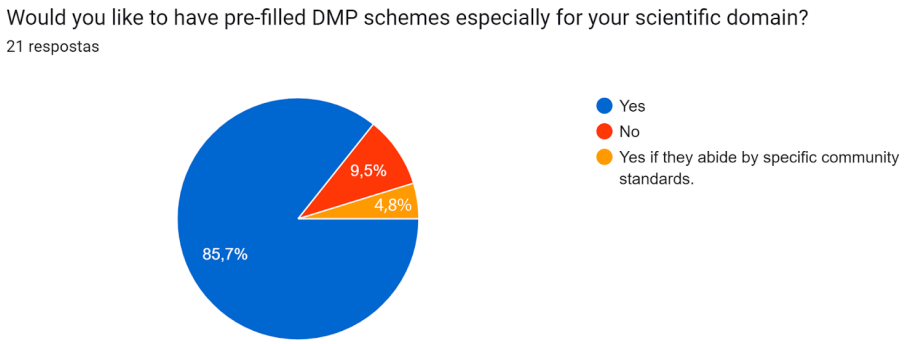

Figure 101: Researchers' answers related to automation of DMP creation process through pre-filled DMP schemes

cally, during the creation of the DMPs, we identified the list of terms for domains history and archaeology, electrical engineering, robotics and agriculture, economic and business, and environmental engineering domain.

For the domain of history and archaeology, important variables related to the designation of the borderland were identified, such as *"belonging to the border, collective identity, inhabitants having concerns that are common to the entire area (a factor that can be detected by historians based on the testimonies left)"*.

For the domain of electrical engineering, specific terms were identified to describe the data providers: *"peak load, voltage, current, power factor profiles, 3-phase and single-phase"*; to describe market agents catalogue: *"generators, retailers, MES aggregators"*.

For the robotics and agriculture project specific terms were defined: to file formats: *"RINEX files; ROSBAG files; RTCM files; ISOBUS files"*; to specific technology used on the sensor: *"3D laser scanner; 3D camera; RGB camera; multi-spectral camera; GNSS solution; vehicle odometry; barometric pressure"*; and to the specific type of the data: *"3D terrain models and map; 3D laser scanner data; system testing data; underlying data"*.

Specific types of data have been defined for the economic and business domain: *"disassembly data; life-cycle product data; product inventories; 3D CAD models; manufacturing planning data; lifespan data; supplier information; bill of materials data; assembly information; subassembly information"*.

For the domain of environmental engineering, specific types of data have been identified, such as *"atmospheric electric field data, gamma radiation data, cosmic radiation data, ocean data, fishing data, meteorological data, cluster ions data"*.

Once the DMP support system has been implemented in researchers' day-to-day work, we will continue to develop Domain Data Protocols for other domains to make it even easier and faster to create quality DMPs in the future.

Moreover, an analysis of the responses related to the automation of the DMP creation process shows that researchers consider it important and necessary (Figure 101). Also, they indicate that having pre-filled DMP schemas, which in turn are based on Domain Data Protocols, will also help to create DMPs easier and faster.

In addition, we received various suggestions from researchers as to what could be automated, according to their point of view. In general, it can be related to the:

- identification of authors through ORCID;
- funder and budget information;
- abstract and responsibilities;
- description of the project, documentation and metadata;
- data preservation issues;
- initial structure of the DMP;
- report structure;

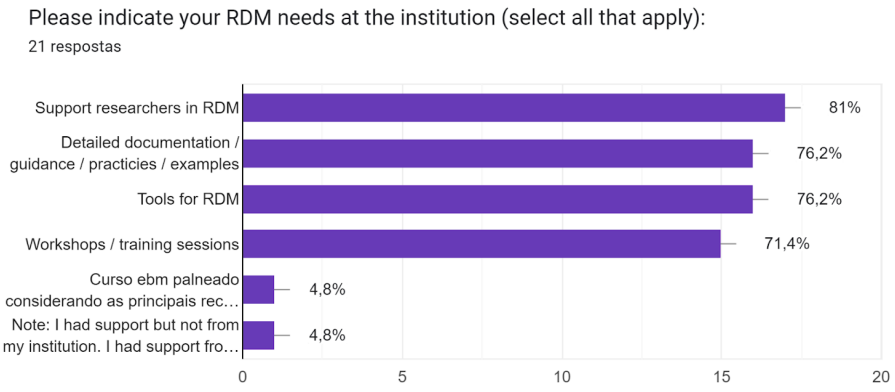

Figure 102: RDM needs at the institution

– data transfer from the project proposal;
– data protection issues.

Analyzing the DMP structure, we also suggest adding scientific domain indication, keywords and bibliographic references in the plan, which could facilitate the reuse of the DMP as an example for other similar cases and be a citable document with the most appropriate structure for publication. In some cases, a glossary may be added.

Some of these suggestions have already been proposed for development in the DMP support system and described in Chapter 6, Section 6.2. Others require more attention and need to be analyzed after the implementation of the DMP support system in the institution.

The last question in our survey focuses on the overall RDM needs in the institution. We want to identify what the researchers need for the execution of the RDM tasks, including DMP creation, and to analyze their difficulties regarding RDM issues in the institution to further improve institutional support.

As it can be seen from the results presented in Figure 102, the majority of researchers (**81**%) want to be supported in RDM activities, **76,2**% want to use detailed documentation, guides, practices, examples and tools to help them execute RDM tasks, and **71,4**% indicate the need for workshops and training sessions (see full researchers' answers in Portuguese and English in Appendix G).

These results highlight that the existence of the data steward at the institution is currently very important. Moreover, they can help in developing and providing all needs required by researchers.

Summarizing, we can state that the application of the same collaborative DMP-building method on the projects from our sample allowed us to obtain detailed results for its systematization, which in turn helps us to propose the DMP support system in compliance with researchers' and institutional needs. The results of this systematization show not only the importance of developing this DMP support system at the institution but also the importance of the existence of a data steward who helps researchers in any matter related to RDM at any point in the project.

The results also show that the collaborative method is quite simple and does not depend on the project budget, project dimension, number of partners and scientific domain; can be used in another institution, either in whole or in part, depending on the specific needs of the institution. For example, they can choose a lightweight support option, leaving only the second scenario, when researchers contact the data steward for support; or, for example, when the data steward does not create the first version of the plan but only analyzes the already prepared one, giving their comments to improve it. In our work, we are developing a more detailed version of

the DMP support system, taking into account different situations, researchers' needs and best practices and recommendations existing in the scientific community. The results show that the data steward supports a sufficiently large number of projects at the same time; however, the choice of using the full version of the DMP support system or its light version, in our opinion, depends on the administration of the institution and the analysis of financial costs.

8.2 THE DMP CONTINUUM – ADDITIONAL PROJECTS

Collaboration with researchers and analysis of the results of our sample show that interest in supporting the creation of DMP is growing, and researchers are proposing new projects where the data steward can help with plans. Moreover, researchers rely on the data steward's help in the future when the DMP support system is already established at the institution. Along with new projects suggested by researchers, there are also projects from previous training work, described in Chapter 5, that are under monitoring. Since the DMP support system is being developed not only to help create plans but also to monitor, improve and keep them as a "living" document, we believe that the analysis of these additional projects helps to confirm the results obtained in the validation sample.

SCRelProg - *Self and Co-regulation in e-Learning of Computer Programming*. This project is related to the development of the pedagogical approach (SimProgramming) and is one of the ten cases described in Chapter 5, Section 5.2.3. It is under monitoring, and the data steward has been accompanying this project since the beginning, guaranteeing support at any time until the end of the project. After the publication of the first version of the DMP, several monitoring sessions occurred, resulting in the creation of the detail DPIA, its revision by the DPO team, publication of the improved version of the DMP on Zenodo [143], and in the creation of the additional documents such as agreements between partners and improved information consent. Monitoring sessions helped to clarify issues about sensitive and personal data management and established new rules for access between members. Our collaboration has been running for forty-one months, showing that the creation of plans is not closed with its publication; the support from the data steward may continue with the project, tracking changes in the project and situations that had not been anticipated. The researchers and the PI of this project value our collaboration, so they sometimes request DMP support for new projects. One such project is "***AVISER***" (scenario 2), where the PI requested assistance in creating a plan during the project proposal stage. The proposal is now awaiting a funding decision. If the project is approved, the data steward will support the researchers in creating a more detailed plan to be submitted within the first six months.

LifeSkillsVR - *Life Skills for Employment in COVID-19 Era through VR Innovation* is another project that was suggested by the same PI during our collaboration (scenario 2). This project is funded by the Erasmus - Programme of the European Union, started in 2021 and intends to take full advantage of VR both as a tool to identify interests and strengths, to develop key life skills required by the labor market and understand which occupation youth can excel and have a bright future[2],[3]. The researcher who collaborated with the data steward in creating the DMP had no experience with RDM, and at the first two meetings the data steward tried to present basic information about RDM in general, the FAIR principles and the processes of creating and monitoring the DMP. Following the collaborative DMP-building method, the data steward created the first draft of the plan and sent it for verification. Moreover, the DMP had to be reviewed by all project members and approved for publication, and the researcher in charge was very active and motivated to get

2 https://www.inesctec.pt/pt/projetos/lifeskillsvr
3 https://lifeskillsvr.com/

all necessary approvals. During the collaboration, the data steward helped to clarify several RDM issues, such as defining the data to be collected during the project, storage and repository information, defining access and backup rules, licensing issues and the use of a Google drive which is not very suitable for storing sensitive data. In addition, the collaboration revealed that the PI had already had contact with the DPO at INESC TEC to review the survey and anonymize the survey responses. Thus, the data steward paid more attention to sensitive and personal data that could be collected during the project and also recommended that collaboration with the DPO should be maintained in case a Data Protection Impact Assessment was needed. As a result of seven months of collaboration, four meetings and researcher interest, the DMP is published on Zenodo [142] and awaits a monitoring session after the project moves on to the next phase, namely the development of the VR system.

SAIL - *Space-Atmosphere-Ocean Interactions in the marine boundary Layer*. This project described in Chapter 5, Section 5.2.4, has been accompanied by the data steward since the beginning of 2019, and the already developed and published DMP is constantly monitored and improved according to the changes that occurred in the project. The collaboration started according to the second scenario when the PI sent an email to the data steward, as they already had various types of support from the data steward for other projects and did not want to start a new project without data steward support and DMP. During the collaboration, the data steward helped with repository definition, data organization and owner issues. During the monitoring sessions, changes in the project were identified, and new versions of the DMP were prepared accordingly and published on Zenodo [15]. Several additional documents were created due to these changes, such as the Technical report on GNSS Post-processing [94], a Technical report on Sensor Data correction [4] and more specific information on datasets, data collection period, tools and naming convention were added. In general, after thirty-one months of collaboration, there are five versions of the DMP [15], and the data steward is always available for the next monitoring sessions, periodically asking the PI about this need. The PI of this project has published several posters at a conference related to the DMP and RDM issues, has used this experience in workshops and promotes RDM knowledge among researchers from their field [14]. Furthermore, the PI continues to show great interest in our collaboration and has invited the data steward to contribute to different proposals. One of them is a new project, **"NewSAT"**, where the data steward will create the first draft DMP to be submitted together with the proposal on the funders' platform.

aMILE - *Application of text mining to clinical reports of patients with acute myeloid leukemia*. This project is also from previous work described in Chapter 5, Section 5.2.7, where the DMP was created according to the second scenario (the PI requested the data steward support). After the publication of the DMP, the PI decided to continue improving the plan, so this project was added to the monitoring list; however, as it was not approved for funding, the improvement of the DMP was stopped in a short time. Despite this, the PI wanted to continue to collaborate with the data steward on other DMP and RDM-related issues. Thus, during the twenty-five months of collaboration, together with the data steward, the PI engaged more with RDM issues and participated in different events such as "RDA hackathon on maDMPs"[4] and "BI Award for Innovation in Healthcare"[5]. Moreover, the PI also participated on the publication of the report "Research Data Management Workflows and maDMPs" [141] and promoted the benefits and importance of the DMP and RDM issues among their scientific domain. Initially, the PI did not have much experience with RDM issues, but our collaboration has improved her skills. The PI continues to expand her knowledge in RDM and DMP topics in-

4 https://github.com/RDA-DMP-Common/hackathon-2020
5 https://biaward.pt/

dependently, clarifying some issues with the data steward, for example, by testing various maDMP-related tools and platforms, such as the Data Stewardship Wizard[6].

Genome Editing - *Facilitated Genome Editing as Responsible Research and Innovation*. This project is one of the ten cases of the previous training work described in Chapter 5, Section 5.2.9, and it aims to improve the efficiency of genome editing using CRISPR by scientifically training eight early-stage researchers. During the first monitor session, the PI decided to publish a revised DMP on Zenodo [212]. It was only one monitor session because the project had come to an end. In general, we can state that, at the beginning of the collaboration, the PI of the project did not have much experience in RDM issues, but with time the knowledge has improved, and motivation has increased. Together with the data steward, they even organized an Open Science workshop on DMP creation for researchers and students of the i3S organization[7]. Thus, the collaboration with the data steward for twelve months resulted in a good training experience where, besides the DMP creation and publication, they expanded RDM knowledge and got more practice.

N-DeciMa - *Neurophysiological bases of decision-making processes*. This project is a part of our collaboration with the Psychology domain at the FPCEUP, where the collaborative DMP-building method was applied together with the metadata model to define controlled vocabularies and Domain Data Protocols (Chapter 5, Section 5.4.2). This project was added to the monitoring list, and on the first monitor session, the DMP was improved and published on Zenodo [139]. During the six-month collaboration, the PI was very motivated, improved RDM-related knowledge and confirmed that the existence of the data steward is very useful. Moreover, the PI confirmed that project members sometimes face the problem of reusing their data created in other projects, so they saw this collaboration as an opportunity to learn how to organize project data, not only for reuse by others but also for their own reuse. Currently, the data steward is waiting for a monitoring session of the plan to add more specific information about the datasets, but this will happen only after the data collection phase of the project is over, so it was decided that the PI would contact the data steward themselves to perform the monitoring session.

VR2CARE - *Multiuser immersive solution for safe group training*. This project is suggested by the PI of the project **"PAFSE"**, described in Chapter 7, according to the second scenario, where the PI requests data steward support. It is funded by the Horizon 2020 project, required DMP creation and aims to develop a unique technological solution that enables people in different physical locations to attend group exercise classes with the active participation of all stakeholders: users, therapists and communities in real-time[8]. So, due to the urgency of the DMP submission, the data steward decided to support researchers on this task. During the three meetings with partners, where the data steward was included from the beginning, the data steward verified that researchers do not have strong DMP and RDM experience, and following the collaborative DMP-building method, created the first draft of the DMP, clarifying all issues related to the RDM. The data steward also established contact with the DPO at INESC TEC due to the existence of personal and sensitive data on the project and guided the creation of some additional documents, such as the informed consent. Researchers were very motivated and grateful to have support on the execution of their deliverable related to the DMP creation, highlighting the importance of having a person, such as a data steward, in the project from the beginning that can answer all questions that arise. In general, collaboration is very active with many email exchanges and telephone contacts, with interest and motivation, given by all members of the project to have a good detailed DMP. After five months of collaboration, the first DMP is created and is under revision by all members of the project to be ready for publishing.

6 https://ds-wizard.org/
7 https://tinyurl.com/yx4ux3ta
8 https://www.vr2care.eu/about/

Table 24: Summary of the additional projects

Project title	Scientific domain	Budget (K euros)	DMP status
SCRelProg: Self and Co-regulation in e-Learning of Computer Programming	Computer and information sciences; other natural sciences; educational sciences	≈ 230	Published
SAIL: Space-Atmosphere-Ocean Interactions in the marine boundary Layer	Computer and information sciences; environmental engineering; earth and related environmental sciences; other engineering and technologies	≈ 100	Published
aMILE: Application of text mining to clinical reports of patients with acute myeloid leukemia	Clinical medcice; other medical sciences; computer and information sciences	Not approved	Published
Genome Editing: Facilitated Genome Editing as Responsible Research and Innovation	Biological sciences; health sciences; medical biotechnology	-	Published
N-DeciMa: Neurophysiological bases of decision-making processes	Psychology; other social sciences	-	Published
NewSAT: Nano-satellite demonstrator to study the ionosphere	Computer and information sciences; environmental engineering; other engineering and technologies	Proposal	Proposal
AVISER: Adaptive Visualization of Intelligently Sensed Extended Realities for training	Computer and information sciences	Proposal	Proposal
LifeSkillsVR: Life Skills for Employment in COVID-19 Era through VR Innovation	Computer and information sciences; economincs and business; educational sciences; other social sciences	≈ 300	Published
VR2CARE: Multiuser immersive solution for safe group training	Clinical medicine; computer and information sciences	≈ 2000	In revision
Submerse: SUBMarine cablEs for ReSearch and Exploration	Computer and information sciences; environmental engineering; other engineering and technologies	-	In revision
DeepField: Deep Learning in Field Robotics: from conceptualization towards implementation	Electrical engineering, electronic engineering, information engineering; computer and information sciences	≈ 800	-
TURING: EITM digiTal Upskilling and ReskIlliNG programme	Computer and information sciences	≈ 245	-
IESMA Summer School	-	-	-
IMMC: Industrial Mobile Manipulator Challenge	Computer and information sciences	≈ 245	-
Green APS: Green Advanced Planning & Scheduling	Computer and information sciences	≈ 250	-
Green MA: Green Manufacturing Accelerator	-	-	-
ConFacts: Multi-layer Connected Factories with hybrid conventional and digital components	Other engineering and technologies	≈ 250	-
Tech2Market: Technology to Market for Competitive Manufacturing in Europe	Other engineering and technologies	≈ 600	-
Demo4Green: Green Manufacturing: Demonstrating technologies to fight Climate Change	Other engineering and technologies	≈ 380	-

Submerse - *SUBMarine cablEs for ReSearch and Exploration*. During the meeting organized with the researchers from **"MYTAG"** and **"inSITE"** projects, described in Chapter 7, they requested support for the creation of the DMP for **"Submerse"** project (scenario 2). The project goal is to create and deliver a pilot activity which would serve as a blueprint for continuous monitoring of many more cables in the future, leading to the opening of new market opportunities and the demonstration of methods to maximize the investments in research infrastructures by using the byproducts of their operations for the purposes of new scientific research. Although this project does not require submission of the DMP, researchers wanted to have this experience as a training session to improve their RDM knowledge and understand how a well-detailed DMP should be created. Within one month, one meeting and several email exchanges, the data steward created the first version of the plan based on an analysis of the existing draft project and sent it to the PI for verification. The DMP is in the revision stage. Moreover, the PI is not sure whether they are going to publish the DMP or not, so, the data steward is waiting for their decision and contact.

DeepField - *Deep Learning in Field Robotics: from conceptualization towards implementation*. This project is suggested by the PI of the **"UNEXUP"** project described in Chapter 7. However, due to a lack of time and the impossibility of supporting this project, the data steward added this request to the list of projects and contacted INESC TEC management, requesting to support this project as soon as possible.

The projects **"TURING"**, **"IESMA Summer School"**, **"IMMC"**, **"Green APS"**, **"GreenMA"**, **"ConFacts"**, **"Tech2Market"**, and **"Demo4Green"** suggested by the PI of the project **"EIT_RIS_Hubs_2021"** described in Chapter 7 are also added to the list of the projects that should be supported as soon as possible. The information about the researchers' interest and requests for DMP creation support is also passed to the INESC TEC management.

Table 24 presents a summary of all additional projects (**19**) from the previous work, which are being monitored or have been proposed during collaborations and meetings with various researchers and PIs. **5** out of **19** projects are taken from the previous work, and their DMPs have already been published. Another **5** projects out of **19** were proposed during the collaboration with the PI and already supported by the data steward; their DMPs are either under revision, created as proposals or published. The remaining **9** projects were also proposed by the PIs during the collaboration but have not received any DMP support yet.

The analysis of these projects shows that the cases using the second scenario increased when the data steward started to collaborate with researchers more closely. In other words, when researchers became aware that they could be supported by the data steward in all DMP-related tasks (e.g., creation, monitoring, improving), the support was requested for more than **14** projects. It shows that the interest in DMP support can also change over time and with the implementation of the well-established DMP support system at the institution. According to these results, the

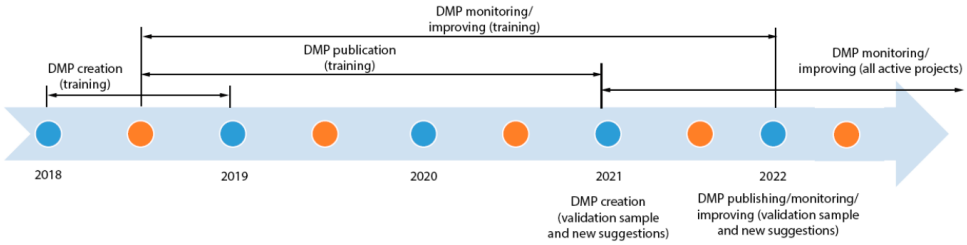

Figure 103: Timeline of DMP-related processes (training projects, validation sample projects and all active projects)

value of metric ***Quant 2 (Number of requests (scenario 2)*** informally is also changing from $\approx$ 3 % to $\approx$ 31 % (15 projects out of 47).

These additional projects also help us to understand the timeline required for the processes of DMP creation, publishing, improving and monitoring. Figure 103 shows that publication of a DMP most often starts after 6 months from the start of its creation, occasionally earlier. Improvement, monitoring and also publication processes can take more than a year or even two. These values thereby confirm our already obtained results in the metric ***Quant 9 (Time)***. It also explains why few DMPs from projects in our sample are published at the moment, most of them are in preparation or in revision. The publication of plans will be in the near future, and any data steward at the institution with the DMP support system could continue to support researchers.

The results of the analysis of the additional projects do not change the results obtained during the systematization of the collaborative DMP-building method applied in the projects in our sample. The additional results only underline the importance of the establishment and implementation of the well-organized DMP support system, and its importance in the long term.

8.3 DISCUSSION AND RECOMMENDATIONS

Summarizing all results obtained from the systematization and application of the collaborative DMP-building method in projects from our defined sample, the answers to the survey, the collaboration between the data steward and the researchers, and the analysis of the additional projects, we can state that all projects are different, but during the support, the data steward support is adapted for each one. One of the contributions of this work is therefore a set of recommendations to the data stewards that can help them support researchers effectively and with more quality in DMP-related processes.

Recommendation 1: The data steward should be flexible and adaptable to the different projects, from different scientific domains, taking into account their specifics.

The data steward has managed to find an approach to every project and researcher, despite the scientific domain, budget, development dynamics, duration of the project, countries and members of the project team.

As we can see, all the results of our analysis and all the data analyzed are interlinked. It is easy to explain because the DMP creation process involves many RDM topics that are also linked with each other. Moreover, the DMP creation process is a time-consuming task, and it is aggravated by the fact that researchers have to think and plan these topics in advance to avoid difficulties and unforeseen situations dur-

ing the project. This is why researchers always indicate that a DMP support system and the existence of a data steward are important in an institution because by getting help with RDM issues, they can guarantee good data management and project compliance with RDM and the funder's requirements.

Recommendation 2: Use both scenarios on the DMP support system for request support (1 - researchers request support from the data steward; 2 - data steward offers support).

According to the results, it can be argued that the two proposed collaboration scenarios, where either the data steward contacts researchers or vice versa, are very similar, but there are still some peculiarities. For example, when the data steward contacted the researchers, they could easily reject collaboration because creating a plan for the project was not mandatory. In the case where researchers contact the data steward, they are more interested and motivated in DMP creation and are more open not only to create the plan but also to its improvement, monitoring, and publishing. The first scenario was used more during the systematization of the collaborative DMP-building method, where the data steward contacted projects and offered DMP support. The second scenario is used more in the data stewards' day-by-day and could be required more frequently when the DMP support system is implemented officially at the institution. That is why we propose using both scenarios on the DMP support system, which fully cover all support needs. We also believe that the number of requests from researchers will increase over time because during the work we observed that the more contacts the data steward establishes, the more requests for support appear.

Recommendation 3: The data steward should use different ways to contact researchers and be available to support them.

The results of the systematization of the collaborative method show that many emails that we sent to PIs were not read, were read without attention or were not interpreted in the right way, so they were unanswered. It could have happened because some emails went to SPAM or were lost in emails received in one day. In cases when the data steward did not receive any reply, the data steward tried other ways of establishing contact.

Recommendation 4: The data steward should disseminate information about the existence of RDM and DMP support and how to request it, motivating researchers to collaborate as early as possible, thereby helping to avoid undesirable situations and wasted time.

According to the results, not all researchers were aware of the existence of DMP support, and in some cases, they created their DMPs without our support. In some cases, the plans required revision or even full re-writing, which increased the time spent on DMP-related processes, devaluing researchers' time spent creating an already existing plan and forcing them to do double work. In other cases the information on the DMPs was incomplete; for example, there was no description of the collection of sensitive or personal data, and, by following, insufficient documents needed to be created to manage these types of data (e.g. informed consent). It was also observed that when a DMP is created at an early stage of the project, it is easier and quicker to create than at a later stage when the time and motivation of the researchers are no longer dedicated to reworking and correcting a large document on which much time and effort has already been spent.

Recommendation 5: The data steward should organize and perform different training sessions for researchers.

Results also show that some projects had no obligation to submit the DMP to the funder, but the PIs were interested in learning how to make these plans in order to understand better what the plan is and what this plan has to include. In other words, to do the training for the following projects, where the plan could be mandatory, or even learning for internal data management of the project. Such interest has a positive impact on the promotion of the DMP support system that is being developed and shows the need for training sessions organized by the data steward at the institution.

Recommendation 6: The data steward should use a complete interview guide covering various important aspects of the DMP-related processes.

In some cases, researchers thought that the plan would be created completely without their involvement and would answer questions they had about the project not related to data management but, for example, with the choice of the best data quality validation method or a better tool to create a project site. Although the data steward can clarify RDM questions and advise suitable tools and methods for many topics related to the data, the data steward does not always have specific technical knowledge of all tools used in the specific scientific domain. The data steward can give advice and orient researchers, but it's the PI who has to make the final decision. The data steward is not part of all the work that occurs in the project, is not usually at meetings between partners, and never has as much detailed information as the PI. In other words, collaboration has to include two sides complementing each other, thus ensuring good quality in the end. This is why collaboration should begin as early as possible so that the data steward can request assistance on any issues from various departments, such as IT, and provide the necessary support in time.

Moreover, these cases led us to make several changes to the interview guide (see Appendix B). In the first version of the interview protocol, the data steward did not emphasize that this work requires a lot of time and effort on the side of the researcher as well. In the improved version of the interview protocol, we added the "Motivation and importance" point where the data steward emphasizes that researchers have to realize that creation of the DMP takes effort and time from their side also because the data steward would never be able to create a good quality plan without them. The data steward can help researchers make their projects FAIR and align with RDM's and funders' requirements by providing the necessary support and tools, but cannot provide answers to all specific project-related questions. For example, the data steward may recommend a more appropriate repository, but the final decision will be made by the PI and other members of the project.

Recommendation 7: The data steward should emphasize that the creation of a DMP is a collaborative effort among all project participants.

There are cases where INESC TEC researchers are not responsible for DMP creation in the project. It was observed that in most of these cases, there is a deliverable related to this task, and the entity responsible for the DMP creation is indicated since the project proposal. Sometimes the researchers say that since INESC TEC is not listed in the deliverable, they do not need to be part of this task. However, since the DMP is desirable to be the unique document for each project and should describe information related to data management from all project members, the data steward underlines the importance of the participation of all members of the project every time. Researchers need to understand that DMP creation is not a task for only one member, but it is collaborative work where all members are responsible. We also believe that the data steward can review the plan and improve it even if it is

created by other members of the project.

Recommendation 8: The data steward should verify what type of support is more appropriate.

In some cases, the researchers contacted other members to propose and discuss data stewards' support. In one option, if the DMP is already created, the data steward offers to analyze the existing DMP and propose improvements. In another option, where DMP does not exist, the data steward can use the same collaborative DMP-building method used for common cases. In both options, responsibility for validation of the plan, authorization of its publication, organization of the meetings between project members to fill the DMP and other types of contacts depending on the project member in charge of DMP creation. This type of collaboration takes more time, more effort, and more email exchanges if it was directly with the researchers from INESC TEC. In other words, for all actions related to the DMP creation, the data steward needs to request authorization through the INESC TEC researcher/PI with the coordinator of another project team.

These situations related to the authorization issues led us to some changes in the interview guide (see Appendix B) and our collaborative method, adding a new step for verification if researchers have DMP created by other project members and proposing, as a consequence, our support. This verification is essential because, as the results show, not all researchers know about documents created by other members indicated as people in charge of deliverables related to DMP. These changes do not modify the essence of the collaborative method but help to improve and organize it to anticipate the complexities associated with DMP creation in larger dimensions projects in advance. For this reason, we recommend that at the beginning of the collaboration, the data steward clarifies which type of support is more appropriate: support only the part of the plan that belongs to the data stewards' institution; support all project partners of any institution and the whole plan; cooperate with data stewards of other institutions.

Recommendation 9: The data steward should clarify who is a project coordinator and who is in charge of DMP creation.

During the work, it has also been noticed that collaboration in the creation of the plan does not always take place with the PI. In some cases, the PI, after the first meeting, connects the data steward with the other researcher of the project more suitable for DMP creation. However, in these cases, the PI is still the main person who must validate the plan and authorize its publication. If INESC TEC is not the project coordinator and the PI does not know if there is any person in charge of DMP creation, it is important to check this. Therefore, we will also add this step for verification of the DMP support system to avoid duplication of work or other difficult situations. This change in the collaborative DMP-building method also does not entail any global modification, only clarifying some points, so does not require new tests.

Recommendation 10: The data steward should explain the benefits of the DMP creation by demonstrating different examples, with different a type of data and for different scientific domains

At the beginning of the collaboration, some researchers had questions about the benefits and usefulness of DMP, and the data steward tried to answer all of them by giving some concrete examples. For example, the data steward asks where five years from now, researchers will find their data in case of project continuation or the necessity of sharing their data with a third party? And if the PI leaves the project, who knows where and how the data is managed? On top of that, the data steward

can also show some examples of plans already created; it works better when these plans are from the same scientific domain. After clarifying issues, researchers opt for DMP creation, even if it is for the better internal organization of data within a project between colleagues.

Another reason why researchers have questions about DMP creation is related to the type of data they collect. They think that the DMP is necessary to be created only for projects that collect data from sensors or some measurements, but not for developed mathematical models, for example. The same doubt also appears when researchers do not collect sensitive or personal data. In these cases, to clarify questions, the data steward uses examples and explains the benefits of data management in the project and, consequently, the creation of DMP. In addition, practice shows that sometimes researchers at the beginning of the project have no idea that they will collect sensitive or personal data, or do not know how to manage them correctly. For example, in some cases, it is not advisable to store personal RAW data obtained from a survey in the same folder as processed data or data encoded using an encryption list. Hence, the data steward offers at least an analysis of the project context and the type of data collected or created during the project. This verification step is also added to our collaborative DMP-building method and, as other changes, doesn't modify the essence of this method. Moreover, the aforementioned questions reveal the need to implement a number of training activities related to DMP and RDM in general. We should promote them within the institution, thereby raising the level of knowledge of researchers. In this case, the data steward could act as a trainer and prepare various materials, webinars, courses and training sessions.

Recommendation 11: The data steward should conduct the monitoring session as planned, even if the researchers indicate that no changes on the project have occurred.

Regarding the monitoring process, in the course of the collaboration, there were cases when it was necessary to monitor the DMP. Although the researchers had a deliverable related to DMP monitoring, in some cases, they stated that it was not necessary because there were no major changes in the course of the work. However, practice shows that there are almost always changes or small questions that may arise during a monitoring session. For example, if at the beginning of the project it was not clearly defined which data repository would be used for datasets publishing, in one of the first monitoring sessions, they could already indicate this. Analysis of these situations shows that the date for monitoring sessions in the DMP support system should be chosen together with the PI and carried out without fail, even if the researchers mention that there are no changes on the project. If there are indeed no changes that can be registered in the DMP, the data steward should formally state this in the DMP, and continue to keep it as a "living" document according to the RDM and funder requirements.

Recommendation 12: The data steward should verify feedback from funders after submission of the DMP for evaluation, if any.

Another step that we consider important to add to improve the collaborative DMP-building method and DMP support system is related to feedback received from funders after a DMP has been submitted. This change is not modifying the essence of this method and does not require new tests, allowing the improvement the DMP after its submission in any monitoring session. Some funders provide feedback and indicate to researchers which part of the DMP needs to be improved. So this verification is added to our collaborative DMP-building method, which is the basis of the DMP support system.

Recommendation 13: The data steward should analyze researchers' DMP even if it has been created without our support.

We have also added a new step to the collaborative DMP-building method, which allows the data steward to check DMPs created without any support. The data steward can then analyse the plan, add comments to improve it, find any inaccuracies or confirm that the plan is of good quality and detail.

Recommendation 14: The data steward should offer to create a DMP-light version, in case it is favourable for the project

In general, it has been observed that the larger the project dimension, budget and number of project participants, the more complex the content of the DMP. Namely, more data management and access rules exist, and more time is spent on creating the DMP as more meetings with partners are required to accept all the established data management rules. Such projects may need to create a light version of the DMP that could be available for publication. During the collaboration with researchers, it was noted that some of these projects already had the first version of the DMP. However, these DMPs were quite large, more like a Project Management Plan. They are for internal use only and can't be published, which in turn makes them difficult to analyze, monitor and improve. Thus, the DMP-light version could be seen as more suitable for these projects for further monitoring, improvement and publication, as advised by RDM and funder requirements. The results show that PIs like this idea, which also led us to include this step in our collaborative method. In our view, this addition does not modify the essence of our collaborative method, only expands the ability to support researchers to have quality DMPs that meet all RDM recommendations. These cases also highlight the importance of DMP support by the data steward from the very beginning to avoid creating large plans and adding unnecessary information to them, saving researchers' time.

The creation of light versions of the plans, in turn, requires a more careful analysis by the data steward. Plans that are not intended for publication but only for internal use may contain private, business information, or sensitive data that cannot be disclosed without the authorization of all project partners.

Recommendation 15: The data steward should create as detailed a first version of the plan as possible.

During the collaboration, we also noted that researchers have more interest, free time and motivation to create the first version of the DMP than its monitoring. Sometimes, the data steward needs to double the effort to perform scheduled monitor sessions for improvements of the already existing DMPs. In our point of view, it is related to the lack of time and the existence of other tasks that researchers need to perform during the project. Therefore, we underline the importance of creating a good plan at an early stage of the project. The support by the data steward at this stage, as we have seen, is obvious, thus creating a good initial structure of the plan, leaving some "open" questions to be filled in during the monitoring sessions.

Recommendation 16: The data steward should collect more information and documentation related to the project in advance.

Finally, we can state that the more information and documentation related to the project researchers share with the data steward from the beginning, the faster and more detailed the data steward can create the first draft of the plan. For example, a lot of information related to the type of data, responsibilities of the project members, data collection method, instruments, and other relevant issues can be found in the detailed description of the Work Packages in the project proposal. That is

why we recommend collecting more information related to the project in advance. We also propose the inclusion of a DMP support system on IRIS, connecting it with the database with project documentation (Chapter 6, Section 6.3, Figure 71). Thus, the data steward could analyze all existing documentation before contacting the researchers, saving them time.

The above analysis presents interesting results but in our opinion, they could be improved if:

1. the DMP creation for projects is mandatory and required by the funders;
2. the DMP creation starts already at the preparation stage of the funding application;
3. the RDM policy is formally adopted and established at the institution;
4. the institutional DMP support system is implemented, and researchers know where to ask for help.

Therefore, given all the results of this work, in the next section, we propose an extended collaborative DMP-building method and DMP support system that includes all recommendations identified during this work.

8.4 PROPOSAL OF THE EXTENDED COLLABORATIVE DMP-BUILDING METHOD

Summarising the results and taking into account all the recommendations identified during the work, in order to extend collaborative DMP-building method we should include the following:

1. add the analysis of the projects in the institutional database that are in the proposal preparation stage and may require creating plans;
2. add a step to verify which of the project partners is in charge of DMP creation to obtain authorization for data steward support;
3. add a step to check whether exist a data steward is involved in another institution;
4. add a step to check the existence of a DMP created either by our institutional project researchers or some other project partner;
5. add a verification step of the type of support that the data steward could provide in case the DMP is created by another partner. This step consists of either:
   - the data steward only supports the part of the plan that belongs to our institution;
   - the data steward supports all project partners and the whole plan;
   - the data steward collaborates with the other data steward in creating the DMP;
6. add a step of analyzing the context of the project and the type of data that will be collected/produced during the project, even if the researchers do not want to create a plan, assuming that a plan is not needed for their data;
7. add a step where the data steward shows various examples of plans, preferably created for the same scientific domain;
8. add a step of appointment of clear dates for monitoring sessions, documenting the information after each session, even if it does not contain any changes;
9. add a verification step related to the feedback from the funders received by the researchers after the DMP has been submitted;
10. add a DMP analysis that was created without data steward support, to improve it or to confirm that the plan is of good quality and detail;
11. add a step to create a DMP-light version in case the plan already created, contains a lot of information (e.g. similar to Project Management Plan), is difficult to monitor and improve, and is not meant to be published;
12. add a step of collaboration with other stakeholders when is necessary.

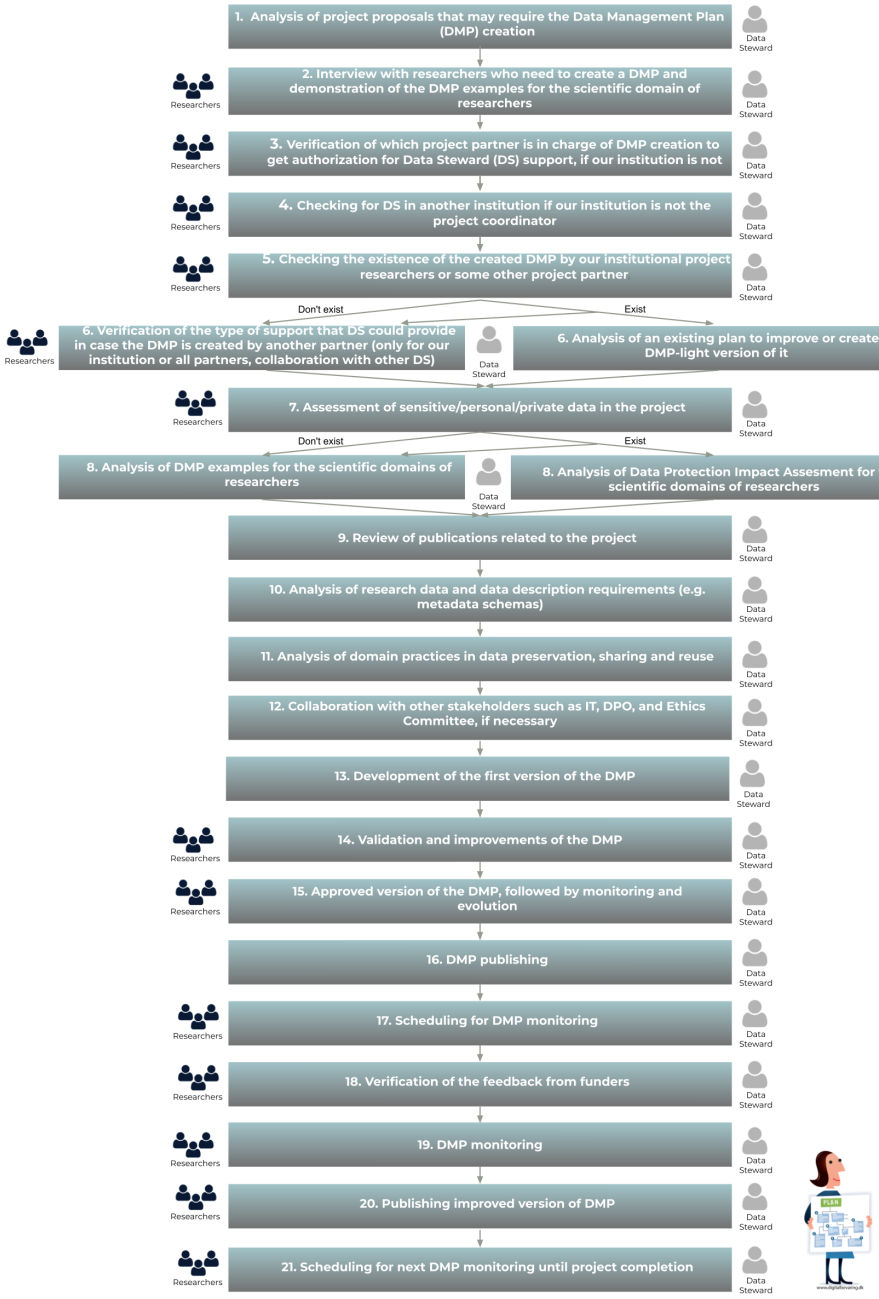

Figure 104: Extended collaborative DMP-building method

In our opinion, all these steps do not change the global essence or structure of our collaborative DMP-building method as a whole, they are not mandatory but advisable. Moreover, we believe that these changes will only improve the collaborative method, making it more flexible and covering most of the difficulties that researchers and data stewards face performing DMP-related tasks. The collaborative method could be used both in its original version (described in Chapter 5, Figure 52) and in its extended version (presented in Figure 104), depending on the needs of institutions and researchers, as well as on the time and resources that an institution can spend on DMP support services. Thus, according to identified points, our extended collaborative DMP-building method includes the following activities.

First of all, the data steward checks whether the researchers need to create a draft DMP to submit with the project proposal (Figure 104, Step 1). After that,

during the interview (Step 2 in Figure 104), the data steward explains the benefits of DMP creation, shows different examples of DMPs, preferably for the same scientific domain and motivates the researchers to collaborate. Moreover, the data steward highlights that researchers will need to take time to create a DMP while the data steward supports them in this task. After that, the data steward gets familiar with the project by analyzing its context and determining the type of data that will be collected. If the researchers believe that it is not necessary to create a plan, the data steward analyzes the reason for this decision. In cases where it is confirmed that there is no need for a DMP, the data steward does not insist on support.

In the next step (Figure 104, Step 3), the data steward needs to check which project partner is responsible for DMP creation. If the DMP creation will not be carried out by our institution, the data steward needs to get authorization for support, for access to all necessary documents and confirm that the other project member actually wants our support. If so, we need to find out whether they have their own data steward to help create the plan (Figure 104, Step 4) and whether any version of the DMP has already been created (Figure 104, Step 5).

In both cases (plan created or not), the data steward needs to specify what type of support will be provided. It could be support for a part of the DMP that belongs to our institution, the whole plan for all project partners or collaboration with another data steward. And if the plan has already been created, the data steward then proceeds with the analysis to improve it, or create a DMP-light version of it (Figure 104, Step 6). If no plan has been created, the data steward will follow steps 7-11, which are similar to our original collaborative DMP-building method. In short, the data steward determines whether there is sensitive, personal or private data, checks whether other additional documents are required, analyzes publications and data related to the project, and analyzes existing description requirements and practices for data preservation, sharing and reusing (Figure 104, Step 7-11). Then, in step 12, the data steward establishes contact with other stakeholders, if necessary. This could be the IT department, the DPO team or the Ethics Committee, among others (Figure 104, Step 12).

The data steward then proceeds to the creation of the DMP draft (Figure 104, Step 13) for subsequent verification and improvement by the researchers (Figure 104, Step 14). Furthermore, the DMP should be approved by all project partners and be authorized for publication (Figure 104, Step 15,16). Scheduling the monitoring process for the DMP is the next step (Figure 104, Step 17). During the preparation for the DMP monitoring, the data steward checks whether the researchers have received any feedback from the funders (Figure 104, Step 18). During the monitoring process, the plan is improved, taking into account all comments and changes that have occurred in the project (Figure 104, Step 19). Then, the improved version of the DMP is published (Figure 104, Step 20), and a new date is chosen for the next monitoring session (Figure 104, Step 21). The monitoring and improvement processes continue until the end of the project.

This method extends the original collaborative DMP-building method, which has already been tested and applied in various projects from different scientific domains. So, since it does not change the essence, the newly added steps do not change the fact that, like the original collaborative method, the extended version can be used in projects with different budgets, with different numbers of project members, with different types of data and at different stages of the project. Moreover, as was mentioned above, the collaborative method can be used step by step, removing unnecessary steps or adding others depending on researchers and institutional needs.

This extended version of the collaborative method is proposed with all the suggestions and results obtained during the development of this work and also can be used as the basis for an institutional DMP support system, which requires an approved institutional RDM policy.

The policy version we propose and describe in Appendix F is developed according to the analysis of the institutional and researchers' needs, RDM and funders'

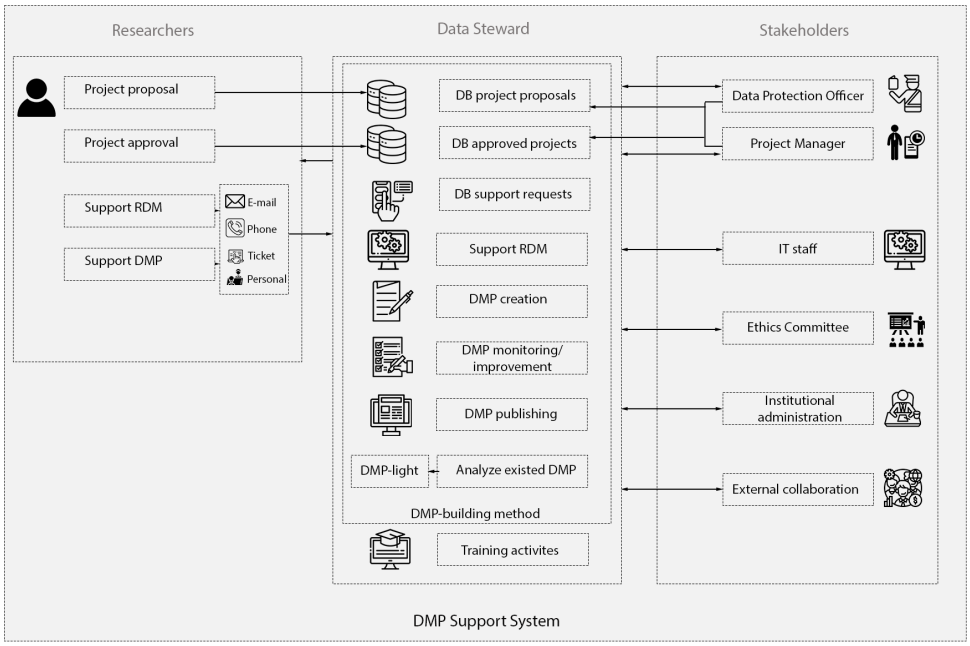

Figure 105: Scheme of the institutional DMP support system

requirements, and based on an analysis of different examples of RDM policies in different institutions. This policy can be approved and implemented in the institution as presented or adapted.

We believe that with the implementation of the institutional RDM policy organizational processes related to RDM support will be improved, researchers will be more involved in RDM, data will be more consistent with FAIR principles, and projects and the institution, in general, will be more compliant with all existing RDM and funder requirements. Moreover, the implementation of the RDM policy will promote the formal implementation of the DMP support system, making it one of the main official points where researchers can get support for all their RDM-related questions.

8.5 INTEGRATION OF THE DMP SUPPORT SYSTEM INTO THE INSTITUTIONAL STRUCTURE

The practical and theoretical work we have done, the application and systematization of our collaborative DMP-building method, and the analysis of the data and the literature allow us to propose a DMP support system that can be implemented in any institution. We believe that the proposed DMP support system will help develop and implement the RDM and DMP workflow and related support services in an institution that is still at the beginning of the development of an institutional RDM infrastructure. It can also enhance and extend the existing RDM infrastructure, providing DMP support at any stage of the project lifecycle. The system will support researchers in creating, monitoring and improving DMPs, helping to make their projects and data compliant with FAIR principles and the existing RDM and funders' requirements.

This system is shown schematically in Figure 105 and comprises three main components: Researchers, Data Steward and Stakeholders, and different stages of the project: proposal preparation and after funding approval. The data steward plays a central role in this system, linking researchers with other stakeholders. Both the original collaborative DMP-building method and its extended version can be used

in this DMP support system.

Proposal preparation stage:

In our version of the DMP support system, it is assumed that when a project proposal is created, it is placed in a database that both the data steward, the project manager and the DPO have access. By accessing this database, the data steward can analyze the project proposal and its related documents and perform the following actions:

1. provide support to researchers during the project proposal creation, specifically the part that describes the RDM issues, if any (scenario 1);
2. provide support to researchers in creating the first draft of the DMP if a funder requests it along with the proposal (scenario 1);
3. provide support to researchers in accompanying their project in all RDM-related matters, even if not required at this stage;
4. request assistance from the project manager for additional information and documentation about the project proposal, if necessary;
5. request assistance from the DPO if sensitive and personal data will be collected and processed in the project and create the first version of the DPIA.

Also, at this stage, the researchers themselves can request a support (scenario 2). Support for DMP creation or other RDM issues can be realized by email, phone, ticket system or in person. If a request is made by the researchers or offered by the data steward, all actions are registered in the database for support requests for further monitoring and control. This control can help to generate the reports required for analysis by the institutional administration when necessary.

The data steward can also request assistance from other stakeholders, for example from the IT department, e.g. to increase the size of the hard disk on the repository server, if necessary. The IT department may also contact the data steward in case of changes to any configurations related to the software or hardware of the RDM infrastructure or to internal procedures such as backup rules.

Collaboration with the Ethics Committee or external institutions is also possible at this stage, for example analyzing the requirements for creating an Ethics Committee authorization, if necessary.

Project approval stage:

When a project is approved for funding, the information and documentations enter the database for approved projects, to which the data steward, the project manager and the DPO also have access. In this case, the data steward, after having analyzed the information, can also offer support (scenario 1). This may consist of the following:

1. support for the creation of the plan;
2. support in other matters related to the RDM;
3. support in the DPIA creation;
4. reviewing the existing DMP, improving it or creating a DMP-light for monitoring and publication;
5. publication of the DMP;
6. monitoring of the DMP.

As in the previous case, researchers can also request support for DMP creation, monitoring, review and improvement, creation of a DMP-light version, and publication at any moment of the project lifecycle.

At this stage, the data steward can also collaborate with other stakeholders to be able to answer any questions that researchers have about RDM issues. In this case, the data steward supports not only the researchers on RDM issues but also other stakeholders. For example, the data steward reduces the workload of the DPO by helping to identify projects which require the creation of the DPIA. Also, by creating the first version of the DPIA, the data steward makes a more complete and quality

document version, which requires less time, cost and resources for its analysis by the DPO.

Training activities are also included in the DMP support system, can be organized and performed by the data steward at any time and aim to increase researchers' and staff's knowledge related to RDM issues, including the DMP-related processes.

The proposed DMP support system covers all stages of the lifecycle of the project, from its preparation as a proposal to its completion, and the lifecycle of the data, from their collection to their reuse. Moreover, during the development of the work, we show that the collaborative DMP-building method (original or extended version) on which the DMP support system is based could be used in any institution, with projects of any budget, and from any scientific domain.

Since the data steward plays an important role in the proposed system, they must be proactive, friendly, persistent, patient and attentive to detail. We believe that by using our work as a guide, any data steward will provide better support for DMPs and RDMs. More projects and researchers will get quality support in advance, more institutional projects will be in line with RDM and funders' requirements, and better evaluation of the institution could get at the government and national levels.

However, in order to be successful, the institutional administration must also make efforts to promote the RDM services, policies, and DMP support system, so that researchers know that the institution is doing everything to help them in all RDM and DMP-related tasks, from proposal preparation until the completion of the project, providing appropriate tools and services.

9 CONCLUSIONS AND FUTURE WORK

As the importance of RDM has begun to grow, all activities related to it have also gained significant value within the scientific community. The actions include not only the development of rules, policies and requirements but also the development of tools, infrastructures, methodologies, systems, services and training sessions.

Both the research project and the data resulting from the project have their own lifecycle. In short, the project starts with the planning stage, and once it is approved, the development stage begins in accordance with the objectives and goals set at the planning stage and ends after they are achieved. For data, the cycle also starts with planning, continues with the collection, processing, and analysis, and ends with storage and, if possible, dissemination for reuse. Thus, the planning stage is one of the main stages of any initiative and contains the thought-out activities necessary to successfully achieve any objective within the allotted time frame.

In our work, we build on the fact that the DMP has become one of the main documents that funders have started requesting for research projects. The DMP is seen as a document that helps researchers in the planning stage and is useful for both the project and the data lifecycle. It can be seen as a document that combines together all aspects related to RDM of a project. This is because a well-created DMP contains all necessary information related to RDM during and after the project. This, in turn, helps institutions support researchers in RDM activities, making projects in line with funders' requirements and FAIR principles, which are also important for the scientific community, by making the data more findable, accessible, interoperable and reusable. During the collaboration with researchers, I assumed the role of a data steward and explained the importance of choosing appropriate repositories, DOI assignment, detailed metadata description, data storage and preservation rules, data openness, licensing, and appropriate access and reuse rules. This collaboration was also important to improve researchers' knowledge and their engagement in the FAIRness of the data.

A well-created DMP requires time, effort, specific knowledge and motivation from researchers. Hence, researchers are very interested in tools, systems and services to help them with time-consuming tasks. According to results presented in Chapter 8, the more support they receive in DMP creation and RDM issues, the more their motivation, interest and awareness in RDM issues grow. That is why institutions are developing RDM workflows and infrastructures, implementing different tools, services and methods to assist researchers in all necessary RDM activities.

In this context, the main focus of this work has been on the analysis of different aspects of RDM aimed at improving the institutional RDM infrastructure and workflow, namely the development and proposal of the DMP support system. The system was supposed to be part of the RDM workflow, be suitable for different types of projects and multiple scientific domains, simplify the process of DMP creation and monitoring, and reduce the time required to complete these tasks. Summing up all the work done, it can be stated that the main objective of the work has been achieved, and the proposed research questions were answered.

Next, we consider them in more detail, starting with RQ1 and RQ4, which focus on the system, its organization and application in different scientific domains.

9.1 DMP SUPPORT FOR MULTIPLE DOMAINS IN THE ORGANISATION

RQ1. How will a DMP support system be incorporated into institutional RDM workflow, becoming a required initial activity of each new research project?

RQ4. Will the proposed DMP support system satisfy the requirements of multiple scientific domains and the funders while keeping track of the data management effort in projects?

During the development of the DMP support system, great emphasis was placed on the importance of providing support at any stage of the project. The DMP support system is based on the collaborative DMP-building method described in Chapter 5 that we applied on different projects from different scientific domains to systematize it and provide a strategy for its implementation into the institutional RDM workflow (Chapter 7). The results of the work described in Chapter 8 show that the DMP support system can help researchers at different stages: in the creation of the DMP for a project proposal or after its approval, accompanying researchers during the lifecycle of the project, and even after the end of the project. Moreover, we show that the DMP support system and the underlying collaborative DMP-building method help researchers in different DMP-related processes such as DMP creation, DMP publication, DMP monitoring, and keeping the DMP as a "living" document.

Although our work shows that a DMP can be created at any stage of a project, it is still preferable to create it early in a project. The results described in Chapter 8 show that the DMP created at the beginning of the project helps not only to organize all strategies of data management during the project but also to avoid misunderstandings and unforeseen situations, planning all the necessary resources in advance. In addition, creating a DMP and its support from the beginning of the project also saves time and effort for both researchers and data stewards. Using the collaborative DMP-building method, any data steward can help researchers to create a well-detailed DMP without unnecessary information, focusing on the most important issues and bottlenecks related to the RDM which may be present in a project. Therefore, the inclusion of the DMP support system as an initial activity in the institutional RDM workflow is a logical and more correct decision, bringing more benefits to both researchers and the data steward and the institute as a whole. Taking into account the results obtained during the work with the DMP support system, we also propose an extended version of the collaborative DMP-building method (Chapter 8, Section 8.4), which includes additional tasks that can help researchers from the very beginning of their research project and cover all stages of the lifecycle of the project.

We also show that from the creation of a DMP to its publication and monitoring, there are many activities that data stewards perform, and these activities require a lot of time, effort, varied cooperation and expertise that result in a good and detailed DMP for researchers (Chapter 8, Figure 103). The data steward organizes meetings, interviews researchers, gets familiar with the project context and specification of the scientific domain, analyzes information related to the project to preview difficulties that can appear during the data management activities, supports researchers in any issues related to the RDM, remembers about of the monitoring sessions, among others. Results show that when a data steward accompanies a project from its planning stage, the accompanying process is easier and the plan includes more details regarding RDM issues of the project. Researchers, in turn, have more possibilities to clarify all RDM issues and make the plan without rush and pressure. Moreover, the more plans researchers create with data steward support, the more aware they become, and less effort will require from data stewards in the future.

In this context, in the DMP support system, we propose to organize data steward access to the database of all prepared proposals (Chapter 8) and to implement a mechanism allowing the data steward to receive notifications during the proposal

creation (Chapter 6, Figure 74). This access and mechanism allow data steward contact with researchers at the beginning of the project's proposal creation, according to scenario **1** described in Chapter 7, Figure 84 and scenario **2** presented on Figure 86.

Regarding the interface of our DMP support system, we proposed our version of its implementation using an example of the INESC TEC research institute (Chapter 6, Figure 71), proposing integration of the DMP support system together with other support services on the IRIS platform. We analyzed the RDM workflows at INESC TEC (Chapter 2, Section 2.7) and stakeholders (Chapter 6, Section 6.1) involved in project management and proposed our version of the workflow, which would include DMP support (Chapter 6, Figure 67).

Also, the application of the collaborative DMP-building method on projects from different institutions, scientific domains, different budgets, and specification shows that the DMP support system takes into account the multiple scientific domains and any type of project. The collaborative method and DMP support system based on it simplifies the DMP-related processes and helps researchers to make their data in line with RDM and funders' requirements. According to researchers' feedback, the data steward helps them with all existing difficulties related to RDM in the project for different types of data such as sensitive, personal, private or public (Chapter 8).

In addition, the data steward can analyze existing DMP, create the DMP-light version of the DMP if it benefits the project, or collaborate with data stewards from the institutions of other project partners, helping them with a plan, promoting its publication even if it is not possible to open and publish project data. In other words, the data steward can propose different types of support, be flexible and adaptable to the needs of researchers, and find different ways of resolving difficulties related to RDM. In Chapter 8, Section 8.3, we propose a series of recommendations to the data stewards, based on our results, that can help them support researchers effectively and with more quality in DMP-related processes.

According to the results and feedback obtained from researchers (Chapter 8), collaborating with the data steward and establishing a DMP support system helps to motivate researchers to create plans, reduces their anxiety about this task, explains the benefits of the RDM and shows that the institution is interested in facilitating all RDM-related tasks that researchers have to perform in their projects, offering them support not only in the tools but also in accompanying the researchers throughout the project. Results of the work also show that using the collaborative DMP-building method and DMP support system (Chapter 8, Section 8.5), it is possible to collaborate with other departments and stakeholders when necessary, and connect them with researchers and accompany this collaboration. The data steward in this collaboration is a central element that can support not only researchers but also stakeholders, such as the DPO, by reducing their efforts related to the preparation and analysis of additional RDM documents required for the project. Moreover, following our approach, the DMP support system also can be implemented in any institution, and any data steward using our work as a guide can support researchers in all DMP-related processes.

9.2 METADATA MODELS AND MONITORIZATION MECHANISMS FOR DMP

The research methodology we proposed and the tasks we set for the development of our work helped us to answer the other research questions. Namely:

RQ2: Can flexible metadata models be devised for each domain with a reasonable effort, taking advantage of the Domain Data Protocols defined by the Science Europe Working Group?

and

RQ3. Based on the preliminary examples of the DMP creation methodology, and the concept, and emerging standards for machine-actionable DMP, how can we devise DMP monitorization mechanisms that enforce compliance of the project results with the decisions taken during DMP creation?

In fact, during the development of our work, in Chapter 5, Section 5.4, we explain the process of creating controlled vocabularies, which can be used to create flexible metadata models and Domain Data Protocols. Collaboration with researchers from the Psychology scientific domain and application of the collaborative DMP-building method allows us to define terms for controlled vocabularies that could be the basis for the proposal of the Domain Data Protocol. The Domain Data Protocol can be used during the DMP creation for specific scientific domains, standartize the DMP, and simplify and automate the DMP creation and monitoring processes (Chapter 3, Section 3.4). Although the development of flexible metadata models, controlled vocabularies and Domain Data Protocols require more profound work in the future, researchers already see them as useful tools that could benefit both researchers and data stewards. According to the results (Chapter 5, Section 5.4), they could help data stewards get familiar with the scientific domain faster, use specific terms during the DMP creation, also serve as a reminder for important topics that need to be described in the DMP, allowing the data steward to make the questions for researchers more specific, suggesting different variants of the terms. This, in turn, helps researchers focus on the project's important details and reduces the data steward's time to gather information to create a DMP with good quality.

During the systematization of the collaborative DMP-building method and the development of the DMP support system, we collaborated with different projects from different scientific domains. During the pre-analysis of the projects from the sample described in Chapter 7, Section 7.2, we identified the scientific domains of each project according to the FOS. This classification helped us to analyze whether researchers use specific terms when creating DMPs in each domain. The results presented in Chapter 8 show that during the collaboration with researchers, there is a possibility to identify specific terms that can be used as controlled vocabularies and then used in the Domain Data Protocols. Responding to our survey, researchers indicated that their scientific domains have specific terms that can be used during the DMP creation process. These specific terms may be in constant use in each project of their particular domain, but may not be useful for other domains. This helped us to identify lists of the specific terms for domains: history and archaeology, electrical engineering, robotics and agriculture, economic and business domain, and environmental engineering (see Chapter 8, Section 8.1). In future work, they can be tested as controlled vocabularies and then used to propose Domain Data Protocols.

During our work, we also propose the DMP workflow integration at the institution (Chapter 6, Section 6.2, 6.3). More specifically, we propose our version of the DMP workflow based on the analysis of the processes related to the project management and according to the machine-actionable DMP concept. The analysis was carried out at INESC TEC using the existing DMP structure based on the collaborative DMP-building method, identifying aspects that needed to be improved regarding the maDMP standard and proposing an improved DMP structure at INESC TEC. Following this, we applied it to the two projects, creating DMPs using the new structure and receiving and analyzing the feedback from researchers. Results show that the new DMP structure is more focused on machine-actionable DMPs, and its use can make the DMP support system more interoperable with other RDM systems. Moreover, the development and implementation of the mechanisms related to the maDMP standard help automate some processes in the DMP support

system, for example, monitoring schedules. In Chapter 6, Section 6.2, we describe our proposal for automation, offering not only a description of the mechanisms but also a visual interface for the DMP support system. At the moment, this is only a proposal since its more detailed development is possible only after the DMP support system has already been implemented into the institutional RDM workflow. Despite this, the proposal can serve as a basis for future work, and during the implementation of the DMP support system at the institution, it can be taken into account. Our proposal of the DMP workflow focuses on the maDMP standard and mechanisms for automation of the DMP creation and monitoring processes.

9.3 DMP SUPPORT SYSTEM AND CERTIFICATION OF THE REPOSITORIES

Regarding the last research question *RQ5: Can the DMP support system include the selection of target repositories while offering an institutional solution certified with the Core Trust Seal?* we show in our work that this is possible. In Chapter 6, Section 6.3 we illustrate a possible way to represent a list of the repositories as controlled vocabularies. This list can be used on the DMP support system and help to automate the DMP creation and monitoring processes. In Figure 73 we show how controlled vocabularies could be implemented in the interface of the DMP support system.

During the application of the collaborative method in different projects, we show that following this method, the data steward always advises a more suitable repository for researchers' needs, whether institutional or other more appropriate (Chapter 7). The results show that almost always, researchers have difficulties in the selection of the repository because of a lack of specific knowledge (Chapter 8). Moreover, during the work, we show that the repository is one of the principal tools that make data compliant with FAIR principles, so the institutions are motivated to the development of their repositories or to use more reliable, trustworthy repositories. Although the certification process for the INESC TEC RDM institutional repository has not been realized yet, in Chapter 2, Section 2.6 we reveal that considerable attention is already given to this topic. We also show that during the collaboration with researchers, the data steward always explains the importance of trustworthy repositories, preferably certified ones. This part of the work is still under development and will be continued in the future.

To summarise, the development of a DMP support system requires substantial effort, time, and analysis of different processes and existing RDM requirements, tools, and recommendations. The system we propose is based on the collaborative DMP-building method, which can be used both in its original version presented in Figure 52 and in its extended version illustrated in Figure 104. All tasks that a data steward may perform to support researchers can be executed sequentially or selectively, depending on institutional needs and research requirements. Our work can be used as a guide for other institutions, following the full guideline or a part of it. However, we believe that the more an institution invests in the RDM infrastructure early on, the sooner they can reap the benefits. The more time and effort a data steward invests in supporting researchers in the DMP creation at the beginning of their project, the better their practice and awareness of RDM issues will be, and the burden on the data steward will gradually diminish. However, as results show, the existence of the data steward at the institution and a DMP support system will continue to be important, as difficulties related to RDM issues may arise at any stage of the project lifecycle.

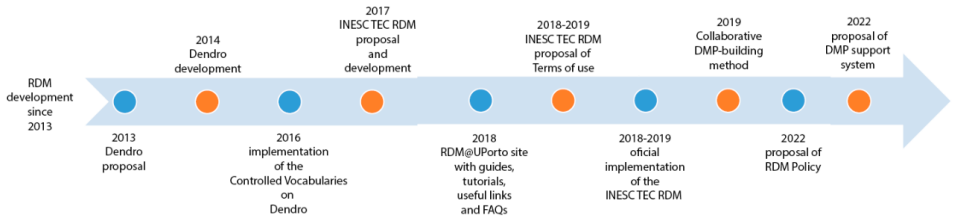

Figure 106: Timeline of the development of the RDM workflow and infrastructure at INESC TEC

9.4 TIMELINE AND EFFORT APPLIED FOR THE DEVELOPMENT OF THE DMP SUPPORT SYSTEM

The development of a DMP support system is a logical step in the development and improvement of the institutional overall RDM infrastructure. The development of the RDM workflow, tools for researchers, services and infrastructure started back in 2013 (see Figure 106).

Although there were many developments related to the RDM during these years, we would like to highlight those that influenced more the DMP support system development, showing the entire chronology of its establishment at the institution. This also shows the time, effort and knowledge required to complete the work because all the RDM tools and developments are based on the analysis of research and institutional needs according to the RDM and funder requirements existing in the scientific community. All tools have also been designed according to the diversity of scientific domains that exist on INESC TEC so that they can always be used for any project of any domain. The following initiatives ate milestones in this work.

2013 - Dendro platform[1] to describe, organize and prepare research data for their deposit and publication on the research data repositories. A large number of experiences were held with researchers from different scientific domains, allowing us to analyze their description difficulties and needs;

2016 - The development of controlled vocabularies and flexible metadata models for the Hydrogen Production scientific domain in Dendro allowed us to describe research data faster and easier with better quality. Moreover, they met the specific description needs of the long-tail projects [133];

2017 - INESC TEC RDM institutional research data repository, data deposit rules and description form have been proposed. Moreover, support for researchers was organized to help both in the description of data and quality control of metadata and control of the deposit process itself. This was the start of creating support for researchers in RDM-related processes;

2018 - The development of the RDM@Uporto website[2] was necessary to disseminate information on tools and services under development and to provide educational, training sessions, different guides and videos aimed at raising the interest, motivation and engagement of researchers in RDM issues. It also helped us to present our work and our team so that researchers know where to look for help when needed;

2018-2019 - The proposal for INESC TEC RDM terms of use and the official implementation of the repository in the institution was realized in 2018-2019, where I, as a data steward, supported researchers using the ticket system[3];

2019 - The development and implementation of the collaborative DMP-building method contributed to the development of support for researchers and the improve-

1 https://github.com/feup-infolab/dendro
2 https://rdm.up.pt/
3 https://ticket.inesctec.pt/

ment of the RDM infrastructure of the institution as a whole. Testing this collaborative method in projects of various domains, as we have seen, served as the basis for the development of the DMP support system;

2022 - RDM policy proposed in this work is the next step towards the formal implementation of the DMP support system.

With these developments, we contribute to the actual research topic, making our work a guide to others in questions of the development and implementing of the well-established DMP support system at the institution, helping others in the formation of decisions. The solution is user-friendly, and the policy can be adaptable for any institution, as the system with all processes and workflows. Moreover, this solution is very easy to implement but depends on the decision of the principal stakeholders and the administration of the institution, taking into account the time, effort, and costs that need to be involved.

Apart from the developments described above, collaboration with different research groups, RDM-related initiatives and organisations working on the development of RDM services and tools, and participation in various conferences, seminars, workshops and training sessions not only allowed us to keep up to date on what was going on during this period but also to contribute to the scientific community and establish different kinds of connections. These activities also helped in the development of this work by providing us with the latest outcomes, different approaches, and best practices existing in the scientific community. In this context, we wanted to highlight the following:

1. Collaboration with the RDA Alliance[4] and groups such as: Vocabulary Services Interest Group (IG), DMP Common Standards Working Group (WG), Active Data Management Plans IG, Metadata Standards Catalog WG, Repository Platforms for Research Data IG, Domain Repositories IG, RDA/WDS Certification of Digital Repositories IG, and Discipline-specific Guidance for Data Management Plans WG, among others.

2. Ongoing presentations of our work and participation in RDA plenaries (one of which was funded by the Early Career Researcher grant): RDA 14 plenary[5],[6]; RDA 16th plenary meeting[7]; RDA 19th Plenary meetings[8].

3. Collaboration with the EUDAT[9]: a European common data infrastructure with integrated services for data preservation and dissemination.

4. MTSR conferences: MTSR 2017[10],[11]; MTSR 2020[12] [136, 145].

5. Different workshops to train researchers.

6. RDA Hackathon on maDMPs[13] where our team "InsTmaDMP"[14] won the third place in all three categories (Innovation, Collaboration and Impact).

7. Participation in national events, meetings, workshops and creation of training courses for the Portuguese community also help us disseminate the results of our work and contribute to the establishment of RDM policies and infrastructures at the institutions: Research Data Alliance Portugal (RDA-pt) node[15], Forum GDI[16], CONFOA[17], and Open Science meetings for trainers[18].

4 https://www.rd-alliance.org/
5 https://www.rd-alliance.org/blogs/rda-fourteen-plenary-early-career-experience.html
6 https://tinyurl.com/nhjpwyc4
7 https://www.rd-alliance.org/rda-16th-plenary-meeting-poster-sessions
8 https://www.rd-alliance.org/rdas-19th-plenary-poster-exhibition
9 https://eudat.eu/
10 http://www.mtsr-conf.org/2017
11 https://www.inesctec.pt/en/news/inesc-tec-researcher-wins-best-student-paper-award-in-estonia
12 http://www.mtsr-conf.org/2020/home
13 https://rda-dmp-common.github.io/hackathon-2020/
14 https://github.com/RDA-DMP-Common/hackathon-2020/blob/master/results.md#instmadmp
15 https://www.rd-alliance.org/groups/rda-portugal
16 https://forumgdi.rcaap.pt/
17 https://confoa.rcaap.pt/
18 https://www.fosteropenscience.eu/trainers-directory

8. Booksprint with experts on RDM requirements and FAIR principles on the development of the training handbook for higher education institutions [80] (see Appendix A for more details).

The basis for our work was the support of researchers from different projects. We collaborated with more than 50 projects from different scientific domains, with different dimensions and different project partners from different countries. Some of these projects were part of the training sample that helped us to define our DMP collaborative method and way of developing a DMP support system (see Chapter 5). Others were part of the validation sample that allowed us to systematize our method in a consistent way, improve it, and expand it (see Chapters 7, 8). We supported researchers in DMP creation and monitoring, data description and deposit processes, which occurred throughout the development of RDM infrastructure in INESC TEC. During this time, ***109*** datasets were deposited at the INESC TEC research data repository and ***10*** - DMPs have been published, not counting those still in development or created without publication.

We also collaborated with different stakeholders and established contact with different institutions. Some of the researchers from other institutions used our collaboration to improve their knowledge and experience as data stewards and are interested in applying our DMP collaborative method and DMP support system to their institutions. So, we can state that this work can be seen as a guide to help any other institutions when planning and developing their RDM workflows and infrastructure.

All the work we have done reflects the importance and relevance of RDM-related topics in the scientific community, which is confirmed by the results. The results also show that researchers started to be more aware of the RDM and funder requirements, and of the FAIR principles required for their data. That is why the institutions began to look for solutions to help researchers perform all RDM activities and to develop support systems. Although the data steward needs to have a proactive attitude, motivation, friendliness, professionalism and perseverance, the development and implementation of the DMP support system require effort and time from all institutional stakeholders.

In our work, we have also seen that DMP creation is a time-consuming and complex task, resulting in a rich source of information relating to RDM in the project. The DMPs created helped us to remember and clarify some of the issues we had during the analysis of the collaborative work that we had with researchers. Researchers also confirmed that the DMPs could be valuable for new members of the project to understand or remember everything that was done during the research, the difficulties researchers encountered and the solutions that helped to solve them. This turns DMPs reusable and a DMP support system beneficial for both the institution and researchers by reducing the time needed to create plans.

9.5 FUTURE WORK

In this context, one of the main aspects of future work is the official recognition, implementation and use of the RDM policy and DMP support system we have proposed, as well as:

1. Developing a method of collaboration with the Ethics Committee that will allow us to better support researchers on ethics-related issues, help them to create the necessary documents, involve third parties when necessary, and monitor this process;

2. Continuing to develop controlled vocabularies and Domain Data Protocols proposals for other scientific domains that will allow us to simplify the creation of DMPs, offer different pre-filled DMP schemes and normalize plans for the same domains, thereby improving their quality and reducing the time and effort involved in their creation;

3. Developing additional training sessions, seminars, and other important guidelines for both researchers and staff of the institution;

4. Continuing and completing the certification process of our institutional INESC TEC RDM repository, which is under review, and creating all necessary documents, policies, and guides; Improving the functionality of the repository, for example, adding statistics and depositing data with restriction or embargo periods;

5. Continuing to automate the DMP creation and monitoring process, according to suggestions of the researchers, e.g. automated collection of the information about researchers in the project using the ORCID database or about the funders and budget, as well as other examples described in Chapter 8;

6. Creation of the institutional DMP template/s that would include bibliographic citations and keywords, which would facilitate the reuse of the DMP as an example for other similar cases and allow it to be a cited document with the most appropriate structure for publication; DMP templates could be different depending on the specification of the scientific domains.

7. Developing a method to track feedback received from funders after the DMP submission in order to improve the plan in time according to the recommendations;

8. Continuing to analyze researchers' and institutional needs, difficulties that exist during the DMP creation and monitoring processes, and improving RDM workflows and DMP support system, when necessary.

Only by bringing all stakeholders together will we achieve better results in all RDM-related tasks, both in the tools and in the entire infrastructure, from proposal preparation to project completion.

REFERENCES

[1] M. Addis. *RDM Workflows and Integrations for Higher Education Institutions Using Hosted Services*. 2015. DOI: 10.6084/m9.figshare.1476832 (cit. on p. 30).

[2] A. Ahmad, J. L. B. Justo, C. Feng, and A. A. Khan. "The impact of controlled vocabularies on requirements engineering activities: a systematic mapping study." In: *Applied Sciences* 10.21 (2020), p. 7749 (cit. on pp. 24, 25).

[3] M. Ahokas, M. E. Kuusniemi, and J. Friman. "The Tuuli Project: Accelerating Data Management Planning in Finnish Research Organisations." In: *International Journal of Digital Curation* 12.2 (2017), pp. 107–115. DOI: 10.2218/ijdc.v12i2.512 (cit. on p. 102).

[4] G. Amaral and N. Dias. *SAIL campaign - Technical report on Sensor Data correction*. Feb. 2021. DOI: 10.5281/zenodo.4518865 (cit. on p. 182).

[5] R. Amorim, J. Castro, J. Rocha, and C. Ribeiro. "A comparison of research data management platforms: architecture, flexible metadata and interoperability." In: *Universal Access in the Information Society* 16:851 (2017), pp. 851–862. ISSN: 1615-5297. DOI: 10.1007/s10209-016-0475-y (cit. on pp. 28, 31).

[6] ANDS. *Guides and Resources: File Formats, Australian National Data Service*. 2017 (cit. on p. 21).

[7] S. B. Ardestani, C. J. Hakansson, E. Laure, I. Livenson, P. Stranak, E. Dima, D. Blommesteijn, and M. Van De Sanden. "B2SHARE: An Open eScience Data Sharing Platform." In: *2015 IEEE 11th International Conference on e-Science*. 2015, pp. 448–453. DOI: 10.1109/eScience.2015.44 (cit. on p. 28).

[8] A. Asher, K. Deards, M. Esteva, M. Halbert, L. Jahnke, C. Jordan, S. Keralis, S. Kulasekaran, W. Moen, S. Stark, et al. "Research data management: Principles, practices, and prospects." In: *Washington, DC: Council on Library and Information Resources* (2013) (cit. on pp. 2, 16).

[9] T. B. Norris. *Data Curation Initiative. University of Miami. 2015-16 Report and Recommendations*. 2016 (cit. on pp. 91, 94, 101).

[10] M. S. Babinec and H. Mercer. "Introduction: Metadata and open access repositories." In: *Cataloging & Classification Quarterly* 47.3-4 (2009), pp. 209–212 (cit. on p. 25).

[11] M. Baca. "Practical issues in applying metadata schemas and controlled vocabularies to cultural heritage information." In: *Cataloging & classification quarterly* 36.3-4 (2003), pp. 47–55 (cit. on p. 23).

[12] A. Bakos, T. Miksa, and A. Rauber. "Research data preservation using process engines and machine-actionable data management plans." In: *International Conference on Theory and Practice of Digital Libraries*. Springer. 2018, pp. 69–80 (cit. on pp. 49, 69).

[13] L. W. Ball. "System effectiveness assurance in response to DOD/NASA requirements (DOD/NASA management practice program ensuring identification, communication and achievement of mission success requirement)." In: *SAE Transactions* 75 (1966), pp. 505–511 (cit. on p. 36).

[14] S. Barbosa and et al. *The SAIL data infrastructure - Open and FAIR research data from the SAIL campaign*. Dec. 2019. DOI: 10.5281/zenodo.3575111 (cit. on p. 182).

[15] S. Barbosa and Y. Karimova. *SAIL Data Management Plan*. Version 2.0.0. Mar. 2021. DOI: 10.5281/zenodo.5070544 (cit. on pp. 38, 39, 42, 111, 182, 226).

[16] C. Bauzer Medeiros, B. Darboux, J. Sánchez, H. Tenkanen, M. Meneghetti, Z. Shinwari, J. Montoya, I. Smith, A. McCray, and K. Vermeir. *IAP input into the UNESCO Open Science Recommendation*. English. July 2020 (cit. on pp. 11, 12, 17, 22, 23).

[17] *Berlin declaration on open access to knowledge in the sciences and humanities*. 2003 (cit. on p. 11).

[18] F. Berman, R. Wilkinson, and J. Wood. "Guest editorial: Building global infrastructure for data sharing and exchange through the Research Data Alliance." In: *D-Lib Magazine* 20.1/2 (2014). DOI: 10.1045/january2014-berman (cit. on p. 19).

[19] F. Berman. *The research data alliance–the first five years*. 2019. DOI: 10.1162/99608f92.5e126552 (cit. on p. 19).

[20] L. Bermudez, E. Montgomery, S. Miller, C. Neiswender, and A. Isenor. "The importance of controlled vocabularies." In: *The MMI Guides: Navigating the World of Marine Metadata* (2011) (cit. on pp. 24, 25).

[21] T. Biro, Y. Karimova, and C. Sales. *Introduction to Research Data Management. Training School of the European Network on Individualized Psychotherapy Treatment of Young People with Mental Disorders (TREATme)*. Version 1.0. COST European Cooperation in Science & Technology (CA16102) Funding: European COMPETE2020 (POCI-01-0145-FEDER-030980) and Portuguese National funds FCT - Fundação para a Ciência e a Tecnologia, I.P (PTDC/PSI-ESP/30980/2017). Oct. 2019. DOI: 10.5281/zenodo.3475149 (cit. on p. 228).

[22] B. W. Bishop, J. Ungvari, H. Gunderman, and H. Moulaison-Sandy. "Data management plan scorecard." In: *Proceedings of the Association for Information Science and Technology* 57.1 (2020), e325 (cit. on pp. 35, 37, 44).

[23] C. L. Borgman. "The conundrum of sharing research data." In: *Journal of the American Society for Information Science and Technology* 63.6 (2012), pp. 1059–1078 (cit. on pp. 8, 9, 16).

[24] C. L. Borgman and P. E. Bourne. "Why it takes a village to manage and share data." In: *CoRR* abs/2109.01694 (2021). URL: https://arxiv.org/abs/2109.01694 (cit. on p. 10).

[25] G. Bot, L. Heller, datawomanHUB, mcancellieri, B. Kramer, T. Ross-Hellauer, ilabastida, helenebr, P. Fernandes, and J. Tennant. *Open Science Training Handbook*. Version 1.1. Apr. 2018. DOI: 10.5281/zenodo.1212538. URL: https://doi.org/10.5281/zenodo.1212538 (cit. on p. 11).

[26] S. Brand, D. Bartlett, M. Farley, M. Fogelson, J. B. Hak, G. Hu, O. D. Montana, J. H. Pierre, J. Proeve, S. Qureshi, A. Shen, P. Stockman, R. Chamberlain, and K. Neff. "A Model Data Management Plan Standard Operating Procedure: Results From the DIA Clinical Data Management Community, Committee on Clinical Data Management Plan." In: *Therapeutic Innovation Regulatory Science* 49.5 (2015), pp. 720–729. ISSN: 2168-4790. DOI: 10.1177/2168479015579520 (cit. on p. 71).

[27] K. Briney, H. Coates, and A. Goben. "Foundational Practices of Research Data Management." In: *Research Ideas and Outcomes* 6 (2020). DOI: 10.3897/rio.6.e56508 (cit. on pp. 8, 14).

[28] K. Briney. *Data Management for Researchers: Organize, maintain and share your data for research success*. Pelagic Publishing Ltd, 2015 (cit. on pp. 8, 9, 11, 14, 15).

[29] I. V. Brvar, P. Doorn, and E. Leenarts. *Presentation. How can Research Data Management Help to Produce Data for Comparative Research?* 2019. URL: https://www.europeansurveyresearch.org/conf2019/uploads/837/465/63/ESRA_2019_DMP_Vipavc_Doorn_Leenarts.pdf (cit. on p. 70).

[30] T. E. Burley and J. D. Peine. *NBII-SAIN Data Management Toolkit*. Report 2009-1170. 2009. DOI: 10.3133/ofr20091170 (cit. on p. 40).

[31] M. Cadman, V. Chavan, S. Patrickson, R. Galt, A. Mader, R. Sood, and T. Hirsch. "GBIF-ICLEI Best Practice Guide for Publishing Biodiversity Data by Local Governments." In: *Copenhagen: Global Biodiversity Information Facility* (2012). URL: http://links.gbif.org/gbif_best_practice_guide_data_publishing_by_local_governments_en_v1 (cit. on p. 109).

[32] P. Caplan. *Metadata fundamentals for all librarians*. American Library Association, 2003 (cit. on p. 23).

[33] P. Caplan. *You call it corn, we call it syntax-independent metadata for document-like objects*. 1995 (cit. on p. 39).

[34] J. Cardoso, L. J. Castro, and T. Miksa. "Interconnecting Systems Using Machine-Actionable Data Management Plans–Hackathon Report." In: *Data Science Journal* 20.1 (2021). DOI: 10.5334/dsj-2021-035 (cit. on p. 69).

[35] J. Cardoso, F. J. Ekaputra, L. Garcia, and C. Jacquemot. *New version of the DMP Common Standard Ontology*. Version 1.0.0. This report is a summary of work done by the Unaturals team at the RDA Hackathon on machine-actionable Data Management Plans that took place between 27th and 29th May 2020. July 2020. DOI: 10.5281/zenodo.3944464 (cit. on p. 69).

[36] J. Cardoso, S. Jones, T. Miksa, A. Hasan, M. Praetzellis, P. Lieby, E. Papadopoulou, and K. Gierend. *Mapping of maDMPs to Funder Templates*. Version 1.0.0. This report is a summary of work done by the TigTag team at the RDA Hackathon on machine- actionable Data Management Plans that took place between 27th and 29th May 2020. July 2020. DOI: 10.5281/zenodo.3944458 (cit. on p. 69).

[37] I. E. Castelli, D. J. Arismendi-Arrieta, A. Bhowmik, I. Cekic-Laskovic, S. Clark, R. Dominko, E. Flores, J. Flowers, K. Ulvskov Frederiksen, J. Friis, et al. "Data Management Plans: the Importance of Data Management in the BIG-MAP Project." In: *Batteries & Supercaps* 4.12 (2021), pp. 1803–1812 (cit. on pp. 43, 44).

[38] J. A. Castro, R. C. Amorim, R. Gattelli, Y. Karimova, J. R. da Silva, and C. Ribeiro. "Involving Data Creators in an Ontology-Based Design Process for Metadata Models." In: *Developing Metadata Application Profiles*. IGI Global, 2017, pp. 181–214. DOI: 10.4018/978-1-5225-2221-8.ch008 (cit. on pp. 23, 24, 26, 227).

[39] C. Clare, M. Cruz, E. Papadopoulou, J. Savage, M. Teperek, Y. Wang, I. Witkowska, and J. Yeomans. *Engaging Researchers with Data Management: The Cookbook*. 2019. DOI: 10.11647/OBP.0185 (cit. on pp. 2, 35, 74).

[40] CODATA. *Open Data and Open Science for the SDGs*. 2017 (cit. on p. 1).

[41] S. Collins, F. Genova, N. Harrower, S. Hodson, S. Jones, L. Laaksonen, D. Mietchen, R. Petrauskaitė, and P. Wittenburg. *Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data*. 2018 (cit. on pp. 18, 22, 23, 29).

[42] E. Comission. *H2020 Programme: Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*. 2017 (cit. on pp. 11, 12).

[43] E. Commission. *Accompanying the document Proposal for a Directive of the European Parliament and of the Council on the re-use of public sector information. SWD/2018/145 final - 2018/0111 (COD). Brussels*. 2018. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=SWD%3A2018%3A145%3AFIN (cit. on pp. 28, 31).

[44] E. Commission. *Directorate-General for Research Innovation. H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020*. 2016 (cit. on pp. 34, 44, 74).

[45] E. Commission. *Horizon 2020. Work Programme 2016 - 2017. Annex L . Conditions related to open access to research data*. 2017. URL: https://ec.europa.eu/research/participants/data/ref/h2020/other/wp/2016-2017/annexes/h2020-wp1617-annex-ga_en.pdf (cit. on pp. 1, 12, 34, 35).

[46] E. Commission and D.-G. for Communication. *The European data strategy : shaping Europe's digital future*. Publications Office, 2020. DOI: 10.2775/645928 (cit. on p. 15).

[47] E. Commission, D.-G. for Research, and Innovation. *Realising the European open science cloud : first report and recommendations of the Commission high level expert group on the European open science cloud*. Publications Office, 2016. DOI: doi/10.2777/940154 (cit. on p. 52).

[48] C. Committee on Data of the International Science Council, C. International Data Policy Committee, C. H. level International Meeting on Open Research Data Policy Practice, CODATA, S. Hodson, M. Barend, U. Paul, and Z. Lili. *The Beijing Declaration on Research Data*. Zenodo. 2019. DOI: http://doi.org/10.5281/zenodo.3552330 (cit. on pp. 1, 34).

[49] T. T. W. Community. *The Turing Way: A handbook for reproducible, ethical and collaborative research*. Version 1.0.1. Nov. 2021. DOI: 10.5281/zenodo.6533831 (cit. on pp. 40, 42).

[50] L. Corti, V. V. den Eynden, L. Bishop, and M. Woollard. *Managing and sharing research data: a guide to good practice*. SAGE Publ. Limited, 2019 (cit. on pp. 1, 9–12, 15–18, 25, 26).

[51] L. Costa and J. R. da Silva. "Dendro: A FAIR, Open-Source Data Sharing Platform." In: *Digital Libraries for Open Knowledge*. Ed. by A. Doucet, A. Isaac, K. Golub, T. Aalberg, and A. Jatowt. Cham: Springer International Publishing, 2019, pp. 384–387 (cit. on p. 26).

[52] E. S. Council. *Open Research Data and Data Management Plans*. 2021 (cit. on p. 16).

[53] E. R. Council. *Guidelines on the implementation of Open Access to scientific publications and research data in projects supported by the European Research Council under Horizon 2020*. 2017 (cit. on p. 11).

[54] N. R. Council. *A question of balance: Private rights and the public interest in scientific and technical databases*. National Academies Press, 2000 (cit. on pp. 9, 10).

[55] A. Cox and E. Verbaan. "Data management planning." In: *Exploring Research Data Management*. 2018. DOI: 10.29085/9781783302802.013 (cit. on pp. 2, 8, 10, 15).

[56] S. J. Cox, A. N. Gonzalez-Beltran, B. Magagna, and M.-C. Marinescu. "Ten simple rules for making a vocabulary FAIR." In: *PLoS computational biology* 17.6 (2021), e1009041 (cit. on pp. 24, 25).

[57] S. Crüwell, J. van Doorn, A. Etz, M. C. Makel, H. Moshontz, J. C. Niebaum, A. Orben, S. Parsons, and M. Schulte-Mecklenbeck. "Seven Easy Steps to Open Science." In: *Zeitschrift für Psychologie* 227.4 (2019), pp. 237–248. DOI: 10.1027/2151-2604/a000387 (cit. on p. 11).

[58] M. Cruz, N. Dintzner, A. Dunning, A. van der Kuil, E. Plomp, M. Teperek, Y. Turkyilmaz-van der Velden, and A. Versteeg. "Policy needs to go hand in hand with practice: The learning and listening approach to data management." In: *Data science journal* 18.1 (2019). DOI: 10.5334/dsj-2019-045 (cit. on pp. 90, 94).

[59] R. Curty, Y. Kim, and J. Qin. "What have scientists planned for data sharing and reuse? A content analysis of NSF awardees' data management plans." In: (2013) (cit. on p. 37).

[60] J. A. Q. Dale L. Goodhue and J. F. Rockart. "Managing the Data Resource: A Contingency Perspective." In: *MIS Quarterly vol. 12 iss. 3* 12 (3 Sept. 1988). DOI: 10.2307/249204 (cit. on p. 36).

[61] J. Davidson, S. Jones, L. Molloy, and U. B. Kejser. "Emerging good practice in managing research data and research information within UK Universities." In: *Procedia computer science* 33 (2014), pp. 215–222 (cit. on pp. 14, 35, 44, 52, 54, 55).

[62] DCC. *Checklist for a data management plan. v.4.0. Edinburgh: Digital Curation Centre*. 2013. DOI: http://www.dcc.ac.uk/resources/data-management-plans (cit. on pp. 47, 53).

[63] DCC. *TheDigitalCuration You-Tube canal*. 2012. URL: https://www.youtube.com/user/TheDigitalCuration/videos (cit. on pp. 53, 54).

[64] B. Department. *Guide for Research Ethics Committee Members*. 2012. URL: https://www.coe.int/en/web/bioethics/guide-for-research-ethics-committees-members (cit. on p. 46).

[65] A. Devaraju, M. Mokrane, L. Cepinskas, R. Huber, P. Herterich, J. de Vries, V. Akerman, H. L'Hours, J. Davidson, and M. Diepenbroek. "From conceptualization to implementation: FAIR assessment of research data objects." In: *Data Science Journal* 20.1 (2021), pp. 1–14 (cit. on p. 21).

[66] D. Dietrich, T. Adamus, A. Miner, and G. Steinhart. "De-mystifying the data management requirements of research funders." In: *Issues in Science and Technology Librarianship* 70.1 (2012) (cit. on p. 34).

[67] I. Dillo. "FAIR Data in Trustworthy Data Repositories." In: *14th International Conference on Digital Preservation. IPRES 2017* (2017) (cit. on pp. 1, 102).

[68] I. Dillo and L. de Leeuw. "CoreTrustSeal." In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 71.1 (2018), pp. 162–170 (cit. on p. 29).

[69] DMPCS. *maDMP-diagram, RDA DMPCS*. 2019. URL: https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard (cit. on pp. 48, 131).

[70] M. Donnelly and S. Jones. *Checklist for a data management plan. Version v.2.2*. 2010 (cit. on p. 53).

[71] M. Donnelly and S. Jones. "DCC Data Management Plan Content Checklist." In: (2009) (cit. on pp. 37–42, 53).

[72] M. Donnelly, S. Jones, and J. W. Pattenden-Fail. "DMP Online: a demonstration of the Digital Curation Centre's web-based tool for creating, maintaining and exporting data management plans." In: *International Conference on Theory and Practice of Digital Libraries*. Springer. 2010, pp. 530–533 (cit. on pp. 42, 53).

[73] P. Doorn. *Science Europe Guidance Document: Presenting a Framework for Discipline-specific Research Data Management*. 2018 (cit. on pp. 2, 5, 34, 35, 48, 70, 123).

[74] P. Doorn. *Science Europe Practical Guide to the International Alignment of Research Data Management*. 2018 (cit. on p. 35).

[75] V. A. Dressler, K. Yeager, and E. Richardson. "Developing a data management consultation service for faculty researchers: A case study from a large Midwestern public university." In: *International Journal of Digital Curation* 14.1 (2019). DOI: https://doi.org/10.2218/ijdc.v14i1.590 (cit. on pp. 2, 74, 102).

[76] DSW. *DSW User Guide*. URL: https://ds-wizard.org/ (cit. on pp. 63, 64, 66).

[77] E. Duval, W. Hodgins, S. Sutton, and S. L. Weibel. "Metadata principles and practicalities." In: *D-lib Magazine* 8.4 (2002), pp. 1–10 (cit. on p. 24).

[78] N. Editorial. "Everyone Needs a Data-management Plan." In: *Nature* 555 (2018), p. 286 (cit. on p. 35).

[79] R. Eichler, C. Giebler, C. Gröger, E. Hoos, H. Schwarz, and B. Mitschang. "Enterprise-Wide Metadata Management: An Industry Case on the Current State and Challenges." In: *Business Information Systems* 1 (July 2021), pp. 269–279. DOI: 10.52825/bis.v1i.47 (cit. on p. 22).

[80] C. Engelhardt, K. Biernacka, A. Coffey, R. Cornet, A. Danciu, Y. Demchenko, S. Downes, C. Erdmann, F. Garbuglia, K. Germer, K. Helbig, M. Hellström, K. Hettne, D. Hibbert, M. Jetten, Y. Karimova, K. Kryger Hansen, M. E. Kusniemi, V. Letizia, V. McCutcheon, B. McGillivray, J. Ostrop, B. Petersen, A. Petrus, S. Reichmann, N. Rettberg, C. Reverté, N. Rochlin, B. Saenen, B. Schmidt, J. Scholten, H. Shanahan, A. Straube, V. Van den Eynden, J. Vandendorpe, S. Venkataram, C. Wiljes, U. Wuttke, J. Yeomans, and B. Zhou. *D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions*. Zenodo, Jan. 2022. DOI: 10.5281/zenodo.6674301 (cit. on pp. 18, 38, 44, 46, 204, 228).

[81] EOSC. *Argos. Details*. URL: https://marketplace.eosc-portal.eu/services/argos/details (cit. on pp. 65–67).

[82] R. Erway. *Starting the Conversation: University-Wide Research Data Management Policy*. ERIC, 2013 (cit. on p. 44).

[83] EUDAT. *EUDAT as a long-term repository for the University of Porto*. EUDAT Data Pilot, available at https://eudat.eu/communities/eudat-as-a-long-term-repository-for-the-university-of-porto. 2017 (cit. on pp. 6, 30).

[84] H. Europe. *Horizon Europe. Programme Guide*. 2022 (cit. on pp. 15, 16, 18, 20).

[85] S. Europe. *Practical guide for the international alignment of research data management. Extended Edition with DMP Evaluation Rubric*. 2021 (cit. on pp. 19, 29, 35, 38, 41, 42, 45, 53).

[86] H. European Commission. *H2020 Online Manual: Open access & Data management* (cit. on pp. 35, 37, 38, 43).

[87] European Organization For Nuclear Research and OpenAIRE. *Zenodo*. en. 2013. DOI: 10.25495/7GXK-RD71. URL: https://www.zenodo.org/ (cit. on p. 26).

[88] "Everyone needs a data-management plan." In: *Nature vol. 555 iss. 7696* 555 (7696 Mar. 2018). DOI: 10.1038/d41586-018-03065-z (cit. on p. 44).

[89] V. V. den Eynden, L. Corti, M. Woollard, L. Bishop, and L. Horton. *Managing and Sharing Data - Best Practice For Researchers*. UK Data Archive, 2011, pp. 1–40. ISBN: 1904059783 (cit. on pp. 1, 8, 12, 14).

[90] B. Faure and K. Bourhy. *maDMP4LS: integration of DMP OPIDoR in the French infrastructure landscape*. 2021. URL: https://www.youtube.com/watch?v=6H6wVNkFGec (cit. on pp. 60–63).

[91] B. Faure, R. Hooft, S. Jones, G. Kakaletris, H. Moa, M. Praetzellis, T. Renner, B. Riley, S. Rust, M. Suchánek, and D. Tziotzios. *Exchange DMPs between DMP Tools using the RDA Common Standard*. Version 1.0.0. This report is a summary of work done by the DMP Exchange team at the RDA Hackathon on machine- actionable Data Management Plans that took place between 27th and 29th May 2020. July 2020. DOI: 10.5281/zenodo.3944308 (cit. on p. 69).

[92] D. Fearon, B. Gunia, S. Lake, B. E. Pralle, and A. L. Sallans. *SPEC Kit 334: Research data management services*. 2013. DOI: 10.29242/spec.334 (cit. on p. 102).

[93] R. Féret and M. Cros. "The embedded research librarian: a project partner." In: *LIBER Quarterly: The Journal of the Association of European Research Libraries* 29.1 (Dec. 2019), pp. 1–20. DOI: 10.18352/lq.10304 (cit. on pp. 89, 94).

[94] A. Ferreira. *SAIL campaign - Technical report on GNSS Post- processing*. Jan. 2021. DOI: 10.5281/zenodo.4447619 (cit. on p. 182).

[95] R. Fidel. "Who needs controlled vocabulary." In: *Special libraries* 83.1 (1992), pp. 1–9 (cit. on p. 24).

[96] I. Foster. "Research infrastructure for the safe analysis of sensitive data." In: *The Annals of the American Academy of Political and Social Science* 675.1 (2018), pp. 102–120. DOI: 10.1177/0002716217742610 (cit. on p. 46).

[97] N. S. Foundation. *Grants.Gov Application Guide A Guide for Preparation and Submission of NSF Applications via Grants.gov. National Science Foundation*. 2011. URL: http://www.nsf.gov/pubs/policydocs/grantsgovguide0111.pdf (cit. on pp. 34, 35, 74).

[98] N. S. Foundation. *Proposal Award Policies Procedures Guide. Chapter II - Proposal Preparation Instructions*. 2019. URL: https://www.nsf.gov/publications/pub_summ.jsp?ods_key=pappg (cit. on p. 104).

[99] E. A. Fox, M. A. Gonçalves, and N. A. Kipp. "Digital libraries." In: *Handbook on Information Technologies for Education and Training* (2002), pp. 623–641 (cit. on p. 145).

[100] M. Getler, D. Sisu, S. Jones, and K. Miller. "DMPonline version 4.0: user-led innovation." In: *International Journal of Digital Curation* 1.9 (2014), pp. 193–219. DOI: 10.2218/ijdc.v9i1.312 (cit. on p. 52).

[101] S. D. Giorgio and P. Ronzino. "PARTHENOS Data Management Plan template for Open Research in Archaeology." In: *2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems Multimedia (VSMM 2018)*. 2018, pp. 1–4. DOI: 10.1109/DigitalHeritage.2018.8810039 (cit. on p. 70).

[102] A. Goben, L. Zilinski, and K. Briney. "Going beyond the data management plan: services and partnerships." In: *The Medical Library Association Guide to Data Management for Librarians. MLA* (2016) (cit. on pp. 44, 45, 102).

[103] A. Government. *Privacy Act 1988, N°119, Compilation N°74*. 1988. URL: https://www.legislation.gov.au/Details/C2016C00979 (cit. on pp. 9, 46).

[104] E. Gowen and J. J. Meier. "Research data management services and strategic planning in libraries today: A longitudinal study." In: *Journal of Librarianship and Scholarly Communication* 8.1 (2020) (cit. on p. 26).

[105] P. Graham. *Managing research data*. 2012 (cit. on pp. 8–10, 13–16, 103).

[106] J. Greenberg. "Metadata and the world wide web." In: *Encyclopedia of library and information science* 3 (2003), pp. 1876–1888 (cit. on p. 22).

[107] J. Greenberg. "Understanding metadata and metadata schemes." In: *Cataloging & classification quarterly* 40.3-4 (2005), pp. 17–36 (cit. on p. 22).

[108] S. Hanrath and E. Radio. "User search terms and controlled subject vocabularies in an institutional repository." In: *Library Hi Tech* (2017) (cit. on pp. 24, 25).

[109] I. Harmon. "Introduction to Research Data Management." In: *2020 Library Immersion Program for Graduate Students* 8 (2020) (cit. on pp. 9, 10, 15).

[110] P. Harpring. *Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works*. Getty Publications, 2013 (cit. on pp. 2, 24, 25).

[111] A. Hasan, A. Fouilloux, and C. Jacquemot. *Integrating maDMP with the data management process*. Version 1.0.0. This report is a summary of work done by the Something team at the RDA Hackathon on machine- actionable Data Management Plans that took place between 27th and 29th May 2020. July 2020. DOI: 10.5281/zenodo.3944434 (cit. on p. 69).

[112] D. Hausen, B. Wham, J. Hrudey, H. Murray, S. Sheridan, I. Anders, Y. Turkyilmaz-van der Velden, and S. Ilamparuthi. *Discipline-Specific Guidance for DMPs WG Case Statement*. 2020. URL: https://tinyurl.com/mrma3yyd (cit. on pp. 71, 72).

[113] N. Health and M. R. Council. *The Australian Code for the Responsible Conduct of Research*. National Health and Medical Research Council, 2018 (cit. on p. 16).

[114] H. Hedden. "Taxonomies and controlled vocabularies best practices for metadata." In: *Journal of Digital Asset Management* 6.5 (2010), pp. 279–284 (cit. on pp. 24, 25).

[115] R. Heery and M. Patel. "Application profiles: mixing and matching metadata schemas." In: *Ariadne* 25 (2000) (cit. on p. 24).

[116] B. P. Heidorn. *Shedding Light on the Dark Data in the Long Tail of Science*. Libr.Trends. 2008 (cit. on p. 1).

[117] D. Hess. " Managing digital asset metadata: The value of flexible metadata model." In: *Journal of Digital Media Management* 3.2 (2014), pp. 130–140.

[118] S. Higgins. "The DCC curation lifecycle model." In: *International Journal of Digital Curation* 3.1 (2008). DOI: 10.2218/ijdc.v3i1.48 (cit. on pp. 13, 14).

[119] L. A. Hook, C. C. Brandt, T. S. Ball, K. N. Fischer, T. L. James, A. J. Kuhaida, D. A. Levine, D. Schmoyer, and W. P. Byrd. *Environmental Data Management Plan for the Environmental Restoration Program*. Tech. rep. Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 1996. DOI: 10.2172/814843 (cit. on p. 71).

[120] M. T. Horsch, S. Chiacchiera, W. L. Cavalcanti, and B. Schembera. "Research Data Infrastructures and Engineering Metadata." In: *Data Technology in Materials Modelling. SpringerBriefs in Applied Sciences and Technology* (2021), pp. 13–30. DOI: 10.1007/978-3-030-68597-3_2.

[121] C. S. Howell Jr. *Supersonic Transport Development Program. Phase 3 Proposal. Boeing Model 2707. Volume 2-3. Aerodynamic Design Report*. Tech. rep. Boeing Co Seattle WA Supersonic Transport DIV, 1966 (cit. on p. 36).

[122] C. R. Hudson Vitale. "The Current State of Meta-Repositories for Data." In: *Curating Research Data, Volume One: Practical Strategies for Your Digital Repository*. Ed. by L. R. Johnston. Association of College and Research Libraries, 2017, pp. 251–261 (cit. on p. 1).

[123] M. Jetten, E. Simons, and J. Rijnders. "The role of CRIS's in the research life cycle. A case study on implementing a FAIR RDM policy at Radboud University, the Netherlands." In: *Procedia Computer Science* 146 (2019). 14th International Conference on Current Research Information Systems, CRIS2018, FAIRness of Research Information, pp. 156–165. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2019.01.090 (cit. on pp. 89, 94, 97).

[124] JISC. *Research data management toolkit, JISC*. 2018. URL: https://www.jisc.ac.uk/full-guide/rdm-toolkit (cit. on p. 14).

[125] S. Jones. *A report on the range of policies required for and related to digital curation*. Tech. rep. Digital Curation Centre, 2009. URL: http://hdl.handle.net/1842/3371.

[126] S. Jones. "Developments in research funder data policy." In: *International Journal of Digital Curation* 7.1 (2012). DOI: 10.2218/ijdc.v7i1.219 (cit. on p. 16).

[127] S. Jones. *How to develop a data management and sharing plan*. 2011 (cit. on pp. 35, 38–42, 44, 47).

[128] S. Jones. *Introduction to Data Management Plans and DMPonline*. 2015. URL: https://www.dcc.ac.uk/events/workshops/introduction-data-management-planning-dmponline (cit. on p. 53).

[129] S. Jones. *Summary of UK research funders' expectations for the content of data management and sharing plans*. 2009 (cit. on pp. 38–42).

[130] E. Kaltman, N. Wardrip-fruin, M. Mastroni, H. Lowood, G. De groat, G. Edwards, M. Barrett, and C. Caldwell. "Implementing controlled vocabularies for computer game platforms and media formats in SKOS." In: *Journal of Library Metadata* 16.1 (2016), pp. 1–22 (cit. on p. 25).

[131] Y. Karimova. *Assessment of metrics for the development of an institutional DMP support system*. Feb. 2023. DOI: 10.25747/YCC7-HH27 (cit. on pp. 5, 141, 144, 150, 151, 166, 167, 169–171, 174).

[132] Y. Karimova. "Flexible metadata models and controlled vocabularies for research data description in multiple domains." In: *Dublin Core 2018 Doctoral Consortium* (2018) (cit. on pp. 6, 24, 227).

[133] Y. Karimova. *MSc dissertation. Vocabulários controlados na descrição de dados de investigação no Dendro*. 2016. URL: https://repositorio-aberto.up.pt/handle/10216/85221 (cit. on pp. 6, 26, 70, 202, 227).

[134] Y. Karimova and J. A. Castro. " Vocabulários controlados na descrição de dados de investigação no Dendro." In: *Cadernos BAD* N.2, (2016), pp. 241–255 (cit. on pp. 25, 26, 227).

[135] Y. Karimova, J. A. Castro, and C. Ribeiro. "Data Deposit in a CKAN Repository: A Dublin Core-Based Simplified Workflow." In: *Digital Libraries: Supporting Open Science*. Ed. by P. Manghi, L. Candela, and G. Silvello. Cham: Springer International Publishing, 2019, pp. 222–235. DOI: 10.1007/978-3-030-11226-4_18 (cit. on pp. 6, 8, 28, 31, 32, 104, 227).

[136] Y. Karimova, J. A. Castro, J. R. da Silva, N. Pereira, and C. Ribeiro. "Promoting Semantic Annotation of Research Data by Their Creators: A Use Case with B2NOTE at the End of the RDM Workflow." In: *Metadata and Semantic Research*. Ed. by E. Garoufallou, S. Virkus, R. Siatri, and D. Koutsomiha. Cham: Springer International Publishing, 2017, pp. 112–122. DOI: 10.1007/978-3-319-70863-8_11 (cit. on pp. 6, 8, 28, 29, 31, 32, 104, 203, 227).

[137] Y. Karimova, J. A. Castro, J. R. D. Silva, N. Pereira, J. Rodrigues, and C. Ribeiro. "Description+ annotation: semantic data publication workflow with Dendro and B2NOTE." In: *International Journal of Metadata, Semantics and Ontologies* 12.4 (2017), pp. 182–194. DOI: 10.1504/IJMSO.2017.093645 (cit. on pp. 6, 8, 30, 227).

[138] Y. Karimova and M. Lopes. *Bibliography and analysis on studies of institutional DMP support services*. Feb. 2023. DOI: 10.25747/YN8F-RA30 (cit. on pp. 76, 78).

[139] Y. Karimova, M. Lopes, C. Botelho, and T. Paiva. *Neurophysiological bases of decision-making processes: dissociating risk and uncertainty in the human brain - Data Management Plan*. Version 1. Work funded by BIAL, grant number PT/FB/BL-2020-252. June 2022. DOI: 10.5281/zenodo.6772779 (cit. on pp. 183, 226).

[140] Y. Karimova, M. Lopes, and I. R. Pinto. *VigilHate - VIGILANT CITIZENS AGAINST HATE: How to counter bystander apathy and increase citizens' commitment against online hate speech?* Version 1. Dec. 2021. DOI: 10.5281/zenodo.5785696 (cit. on pp. 120, 226).

[141] Y. Karimova, M. Lopes, S. Ramos, R. Rb-Silva, S. Barbosa, Â. Lomba, and M. Potterbusch. *Research Data Management Workflows and maDMPs*. Version 1.0.0. This report is a summary of work done by the InsTmaDMP team at the RDA Hackathon on machine- actionable Data Management Plans that took place between 27th and 29th May 2020. July 2020. DOI: 10.5281/zenodo.3944468 (cit. on pp. 7, 69, 125, 129, 182, 227).

[142] Y. Karimova, M. Lopes, and D. L. Silva. *Life Skills VR: Life Skills for Employment in COVID-19 Era through VR Innovation*. Version 1. Apr. 2022. DOI: 10.5281/zenodo.6484756 (cit. on pp. 182, 226).

[143] Y. Karimova, D. Pedrosa, J. Cravino, and L. Morgado. *SCReLProg (Self and Co-regulation in e-Learning of Computer Programming)*. Version 1. Dec. 2021. DOI: 10.5281/zenodo.5750127 (cit. on pp. 38–40, 43, 181, 226).

[144] Y. Karimova and C. Ribeiro. *The collaborative method between curators and researchers in the preparation of a Data Management Plan and Privacy Impact Assessment*. 5 Forum Gestão de Dados de Investigação, Aveiro. 2019. URL: https://forumgdi.rcaap.pt/wp-content/uploads/2019/12/09_5ForumGDI__Yulia_Karimova.pdf (cit. on p. 105).

[145] Y. Karimova, C. Ribeiro, and G. David. "Institutional support for data management plans: case studies for a systematic approach." In: *International Journal of Metadata, Semantics and Ontologies* 15.3 (2021), pp. 178–191 (cit. on pp. 31, 37, 40–46, 74, 90, 104, 141, 145, 173, 203, 228).

[146] Y. Karimova, C. Ribeiro, and G. David. "Institutional Support for Data Management Plans: Five Case Studies." In: *Research Conference on Metadata and Semantics Research*. Springer. 2020, pp. 308–319. DOI: 10.1007/978-3-030-71903-6_29 (cit. on pp. 6, 43, 44, 88, 94, 98, 104, 141, 227).

[147] Y. Karimova, G. Silva, and A. Costa. *FronTowns: Think big on small frontier towns: Alto Alentejo and Alta Extremadura leonesa (13th - 16th centuries)*. Version 1. June 2022. DOI: 10.5281/zenodo.6659636 (cit. on pp. 152, 226).

[148] J. Kim, E. R. Schuler, and A. Pechenina. "Predictors of data sharing and reuse behavior in academic communities." In: *Knowledge Discovery and Data Design Innovation: Proceedings of the International Conference on Knowledge Management (ICKM 2017)*. World Scientific. 2018, pp. 1–25 (cit. on p. 13).

[149] J. Klar, O. Michaelis, J. Straka, U. Wuttke, and R. Rothe. *maDMP export from RDMO*. Version 1.0.0. This report is a summary of work done by the RDMO team at the RDA Hackathon on machine-actionable Data Management Plans that took place between 27th and 29th May 2020. July 2020. DOI: 10.5281/zenodo.3944448 (cit. on p. 69).

[150] A.-L. Lamprecht, L. Garcia, M. Kuzak, C. Martinez, R. Arcila, E. Martin Del Pico, V. Dominguez Del Angel, S. Van De Sandt, J. Ison, P. A. Martinez, et al. "Towards FAIR principles for research software." In: *Data Science* 3.1 (2020), pp. 37–59 (cit. on p. 18).

[151] A. Latif. *EUDAT: Research data infrastructure and European Open Science Cloud vision*. Team ZBW Mediatalk, available at https://www.zbw-mediatalk.eu/en/2017/05/eudat-research-data-infrastructure-and-european-open-science-cloud-vision/. 2017 (cit. on pp. 6, 28).

[152] Y. Le Franc. *Organise, retrieve and aggregate data using annotations with B2NOTE*. EUDAT webinar, available at https://eudat.eu/events/webinar/eudat-webinar-organise-retrieve-and-aggregate-data-using-annotations-with-b2note. 2017 (cit. on p. 28).

[153] D. J. Lee and B. Stvilia. "Practices of research data curation in institutional repositories: A qualitative view from repository staff." In: *PLoS ONE* 12.3 (2017), pp. 1–44. DOI: 10.1371/journal.pone.0173987 (cit. on p. 3).

[154] A. Lefebvre, B. Bakhtiari, and M. Spruit. "Exploring research data management planning challenges in practice." In: *It-Information Technology* 62.1 (2020), pp. 29–37 (cit. on pp. 14, 35–37, 45, 67, 69, 87, 88).

[155] T.-P. Liang. "Integrating model management with data management in decision support systems." In: *Decision Support Systems* 1.3 (1985), pp. 221–232. ISSN: 0167-9236. DOI: https://doi.org/10.1016/0167-9236(85)90241-6 (cit. on p. 36).

[156] M. Libraries. *Strategic Plan for Research Data Services at the University of Minnesota Libraries*. 2019. URL: https://hdl.handle.net/11299/203926 (cit. on pp. 92, 94).

[157] G. Lise, L. R. Xavier, E. Hilde, B. Bradley, B. Claire, B. Dimitri, C. Maxime, D. Rowena, H. Harri, H. André, J. Cécile, L. Tina, R. Gaby, and U. Judit. *Guidance document for scientists on data management, open data, and the production of Data Management Plans. BiodivERsA report*. Sept. 2019. DOI: 10.5281/zenodo.3448251 (cit. on pp. 11, 29, 35, 37–45).

[158] A. Lomba, F. Moreira, S. Klimek, R. HG Jongman, J. P. Honrado, T. Pinto-Correia, T. Plieninger, and D. I. McCracken. "Back to the future: rethinking socioecological systems underlying high nature value farmlands." In: *Frontiers in Ecology and the Environment* 18.1 (2020), pp. 36–42. DOI: 10.1002/fee.2116 (cit. on pp. 109, 132).

[159] M. Lopes. *Metadata model with controlled vocabularies proposed for the Social Sciences domain*. Oct. 2022. DOI: 10.5281/zenodo.7264856 (cit. on pp. 118, 119).

[160] M. L. Lopes C. *MSc dissertation. Descrição de dados de investigação: requisitos de investigadores para modelos de metadados na Psicologia e Ciências da Educação*. 2019. URL: https://hdl.handle.net/10216/122418 (cit. on p. 118).

[161] P. Lord, A. MacDonald, R. Sinnott, D. Ecklund, M. Westhead, and A. Jones. *Large-Scale Data Sharing in the Life Sciences: Data Standards, Incentives, Barriers and Funding Models (The Joint Data Standards Study)*. 2005 (cit. on p. 35).

[162] B. F. Lóscio, C. Burle, and N. Calegari. Data on the Web Best Practices, available at https://www.w3.org/TR/dwbp/. 2017 (cit. on p. 1).

[163] J. M. Ray. *Research data management: Practical strategies for information professionals*. 2013. DOI: 10.2307/j.ctt6wq34t (cit. on pp. 8–10, 13, 14, 22, 25, 103).

[164] T. Machacova and F. Jana. *Friday Coffee 5: Data Management with Data Stewardship Wizard*. 2021. URL: https://www.youtube.com/watch?v=2YGrVGa9JYY&ab_channel=ELIXIRCzechRepublic (cit. on p. 65).

[165] *Machine-actionability. GO FAIR glossary*. URL: https://www.go-fair.org/resources/glossary/ (cit. on p. 47).

[166] *Managing Data@ Melbourne Working Group. Managing Data@ Melbourne: an online research data management training program*. 2017. URL: https://minerva-access.unimelb.edu.au/handle/11343/225138 (cit. on p. 103).

[167] K. Mary Anne. *Managing research data. Research Methods*. 2018. DOI: 10.1016/b978-0-08-102220-7.00021-2 (cit. on p. 103).

[168] M. S. Mayernik. *Metadata Realities for Cyberinfrastructure: Data Authors as Metadata Creators*. Social Science Research Network. 2011. DOI: 10.2139/ssrn.2042653 (cit. on p. 30).

[169] D. T. McAlister. "The Project Management Plan: Improving Team Process and Performance." In: *Marketing Education Review* 16.1 (2006), pp. 97–103. DOI: 10.1080/10528008.2006.11488946 (cit. on p. 115).

[170] W. K. Michener. "Ten simple rules for creating a good data management plan." In: *PLoS computational biology* 11.10 (2015), e1004525 (cit. on pp. 1, 34, 37–42, 44, 45).

[171] T. Miksa, J. Cardoso, and J. Borbinha. "Framing the scope of the common data model for machine-actionable data management plans." In: *2018 IEEE International Conference on Big Data (Big Data)*. IEEE. 2018, pp. 2733–2742. DOI: 10.1109/BigData.2018.8622618 (cit. on pp. 47, 50, 52, 60–62, 68, 69).

[172] T. Miksa, P. Neish, P. Walk, and A. Rauber. "Defining requirements for machine-actionable data management plans." In: *iPRES*. 2018 (cit. on pp. 47–49, 52, 53, 129).

[173] T. Miksa, S. Oblasser, and A. Rauber. "Automating Research Data Management Using Machine-Actionable Data Management Plans." In: *ACM Transactions on Management Information Systems (TMIS)* 13.2 (2021), pp. 1–22. DOI: 10.1145/3490396 (cit. on pp. 47–49).

[174] T. Miksa, A. Rauber, R. Ganguly, and P. Budroni. "Information integration for machine actionable data management plans." In: *International Journal of Digital Curation* 12 (Sept. 2017), p. 22. DOI: 10.2218/ijdc.v12i1.529 (cit. on pp. 35, 42, 44, 47, 48, 50, 52).

[175] T. Miksa, S. Simms, D. Mietchen, and S. Jones. "Ten principles for machine-actionable data management plans." In: *PLoS computational biology* 15.3 (2019). DOI: 10.1371/journal.pcbi.1006750 (cit. on pp. 2, 47–50, 69).

[176] T. Miksa, P. Walk, and P. Neish. *RDA DMP Common Standard for Machine-actionable Data Management Plans*. 2020. DOI: https://doi.org/10.15497/rda00039 (cit. on pp. 2, 47, 48).

[177] T. Miksa, P. Walk, P. Neish, S. Oblasser, M. HOLLAND, T. Renner, M.-C. Jacquemot-Perbal, J. Cardoso, T. Kvamme, M. Praetzellis, et al. "Application Profile for Machine-Actionable Data Management Plans." In: *Data Science Journal* 20.1 (2021) (cit. on pp. 47–50, 52, 53, 55, 63–67, 69).

[178] T. Milotic. *Good enough practices in data management and how to translate these in data management plans (DMPs)*. Sept. 2018. DOI: 10.5281/zenodo.1421739 (cit. on pp. 34, 38–45).

[179] W. H. Mischo, M. C. Schlembach, and M. N. O'Donnell. "An analysis of data management plans in University of Illinois National Science Foundation grant proposals." In: *Journal of eScience Librarianship* 3.1 (2014), pp. 31–43 (cit. on pp. 37, 71).

[180] E. Mitchell. *Metadata Standards and Web Services in Libraries, Archives, and Museums: An Active Learning Resource: An Active Learning Resource*. ABC-CLIO, 2015 (cit. on p. 23).

[181] H. Moa, A. Hasan, and J. Philipson. *EasyDMP and RDA DMP Common Standard*. Version 1.0.0. This report is a summary of work done by the Datatypists team at the RDA Hackathon on machine- actionable Data Management Plans that took place between 27th and 29th May 2020. July 2020. DOI: 10.5281/zenodo.3944275 (cit. on p. 69).

[182] P. S. R. Morton. "A storage and retrieval plan for data on head and neck cancer patients." In: *Journal of Microcomputer Applications vol. 7 iss. 3-4* 7 (3-4 July 1984). DOI: 10.1016/0745-7138(84)90076-9 (cit. on p. 36).

[183] M. Mosley, M. Brackett, S. Earley, and D. Henderson. *The DAMA Guide to The Data Management Body of Knowledge*. 2009. DOI: 10.1161/CIRCULATIONAHA.108.834176.

[184] C. L. Moss Jr. "The Information Systems Plan and the Data Base Plan." In: *Information System Management* 1.3 (1984), pp. 20–26 (cit. on p. 36).

[185] P. Murray-Rust. "Open Data in Science." In: *Serials Review* 34:1 (2008), pp. 52–64. DOI: 10.1080/00987913.2008.10765152.

[186] E. National Academies of Sciences, Medicine, et al. *Life-Cycle Decisions for Biomedical Data: The Challenge of Forecasting Costs*. National Academies Press, 2020 (cit. on pp. 34, 35).

[187] H. Nie, P. Luo, and P. Fu. "Research Data Management Implementation at Peking University Library: Foster and Promote Open Science and Open Data." In: *Data Intelligence* 3.1 (Feb. 2021), pp. 189–204. DOI: 10.1162/dint_a_00088 (cit. on pp. 87, 94, 99).

[188] D. A. Nitecki and A. Alter. "Leading FAIR adoption across the institution: a collaboration between an academic library and a technology provider." In: *Data Science Journal* 20.1 (2021). DOI: 10.5334/dsj-2021-006 (cit. on pp. 86, 98).

[189] S. Oblasser. *Machine-actionable DMP application (DMap)*. The slides are based on a live demo given at RDA 14th Plenary taking place in Helsinki between 22-25 October 2019. Further session material can be found here: https://www.rd-alliance.org/p14 -helsinki-slides. Oct. 2019. DOI: 10.5281/zenodo.3522247 (cit. on pp. 49–51).

[190] OECD. *OECD Principles and Guidelines for Access to Research Data from Public Funding*. 2007, p. 49 (cit. on pp. 11, 12, 15, 16, 18, 26).

[191] C. M. Olson. "The Development and Implementation of a Remedial Investigation Work Plan and Data Management System." In: *Groundwater Monitoring Remediation vol. 11 iss. 2* 11 (2 May 1991). DOI: 10.1111/j.1745-6592.1991.tb00376.x (cit. on p. 36).

[192] OpenAIRE. *Argos. About*. URL: https://argos.openaire.eu/splash/about/how-it-works.html (cit. on pp. 65–67).

[193] OpenAIRE. *OpenAIRE. FAQ*. URL: https://www.openaire.eu/faqs (cit. on p. 45).

[194] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher. "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews." In: *BMJ* 372 (2021). DOI: 10.1136/bmj.n71 (cit. on pp. 74, 75).

[195] E. Papadopoulou. *Implementing the DMP Common Standard on the Argos tool for maDMPs*. URL: https://www.rd-alliance.org/implementing-dmp-common-standard-argos-tool-madmps (cit. on pp. 65–67).

[196] E. Papadopoulou. *Presentation. Argos. Plan and follow your data*. 2021. URL: https://www.um.edu.mt/library/oar/bitstream/123456789/70269/1/ARGOS_plan_and_follow_your_data_2021.pdf (cit. on pp. 65–67).

[197] E. Papadopoulou, A. Bardi, G. Kakaletris, D. Tziotzios, P. Manghi, and N. Manola. "Data Management Plans and Linked Open Data: exploiting machine actionable data management plans through Open Science Graphs." In: *DaMaLOS*. 2020, pp. 1–12 (cit. on pp. 65–68).

[198] E. Papadopoulou, G. Kakaletris, D. Tziotzios, H. Moa, and A. Hasan. *ARGOS: a collaborative tool to plan and follow your data*. June 2020. DOI: 10.5281/zenodo.3898249 (cit. on pp. 66, 67).

[199] S. W. Parham, J. Carlson, P. Hswe, B. Westra, and A. L. Whitmire. *Using Data Management Plans to Explore Variability in Research Data Management Practices Across Domains*. International Journal of Digital Curation. 2016 (cit. on pp. 2, 44).

[200] J.-r. Park and E. Childress. "Dublin Core metadata semantics: An analysis of the perspectives of information professionals." In: *Journal of Information Science* 35.6 (2009), pp. 727–739 (cit. on p. 23).

[201] E. Parliament. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)*. 2016. URL: http://data.europa.eu/eli/reg/2016/679/oj (cit. on pp. 46, 106).

[202] A. S. Petrus and M. Töwe. "Forschungsdatenmanagement an der ETH Zürich: Ansätze und Wirkung." In: *Bibliothek Forschung und Praxis* 43.1 (2019), pp. 49–60. DOI: doi:10.1515/bfp-2019-2002 (cit. on pp. 85, 94).

[203] D. Pilat and Y. Fukasaku. "OECD principles and guidelines for access to research data from public funding." In: *Data Science Journal* 6 (2007), OD4–OD11 (cit. on p. 11).

[204] M. Policy. *Guidance on sharing of research data from population and patient studies. MRC, 2011*. URL: https://epi-meta.mrc-epid.cam.ac.uk/downloads/MRCpolicyguidanceDataSharingPopPatientStudies_01-00.pdf (cit. on p. 35).

[205] I. Popescu. *EOSC-hub: integrated services for the European Open Science Cloud*. 2018. URL: https://www.egi.eu/news/eosc-hub-integrated-services-for-the-european-open-science-cloud/.

[206] L. Pouchard. "Revisiting the data lifecycle with big data curation." In: *International Journal of Digital Curation* 10.2 (2015) (cit. on p. 14).

[207] P. Príncipe, M. Karstensen, M. Loesch, and L. H. Nielsen. "OpenAIRE Guidelines for Data Archive , Literature Repository and CRIS Managers." In: *Proceedings of the DCIM International Conference on Dublin Core and Metadata Applications*. 2013, pp. 203–205.

[208] P. Prıncipe, P. Moura, A. Vieira, and F. Pereira. *Relatório técnico sobre ferramentas para a elaboração de Planos de Gestão de Dados*. 2020. DOI: 10.21814/1822.67530 (cit. on pp. 63, 65–68).

[209] S. Purba. *Handbook of Data Management: 1999 Edition*. 1st ed. Auerbach Publications, 1998 (cit. on p. 37).

[210] J. Qin, A. Ball, and J. Greenberg. "Functional and Architectural Requirements for Metadata: Supporting Discovery and Management of Scientific Data." In: *Proceedings of the DCIM International Conference on Dublin Core and Metadata Applications*. 2012, pp. 62–71. URL: http://dcpapers.dublincore.org/pubs/article/view/3660 (cit. on pp. 22, 39).

[211] C. Ramjoué. "Towards open science: The vision of the European Commission." In: *Information Services & Use* 35.3 (2015), pp. 167–170 (cit. on p. 10).

[212] P. Ramos, Y. Karimova, and A. Olsson. *Facilitated Genome Editing as Responsible Research and Innovation*. Version 1. Oct. 2021. DOI: 10.5281/zenodo.5600183 (cit. on pp. 42, 183, 226).

[213] RDA. *FAIR Data Maturity Model: specification and guidelines*. RDA FAIR Data Maturity Model Working Group. 2020. DOI: 10.15497/RDA00045.

[214] RDA. *RDA for the Sustainable Development Goals. Introduction: Fit with the overall RDA vision and mission*. 2019. URL: https://www.rd-alliance.org/groups/rda-sustainable-development-goals (cit. on p. 1).

[215] RDA18. *WG Discipline specific Guidance for Data Management Plans: WG Discipline specific Guidance DMPs*. 2021. URL: https://www.youtube.com/watch?v=_WaZOqyDgcM&ab_channel=RDAResearchDataAlliance (cit. on p. 72).

[216] RDA18. *WG DMP Common Standards WG: Maintenance and Adoptions of the Common Standard for machine-actionable DMPs*. 2021. URL: https://www.youtube.com/watch?v=6H6wVNkFGec (cit. on p. 69).

[217] N. S. Redkina. "Current Trends in Research Data Management." In: *Scientific and Technical Information Processing* 46 (2019), pp. 53–58. DOI: 10.3103/S0147688219020035 (cit. on pp. 93, 94).

[218] T. Renner. *Turbo-charging Data Management Plans*. Feb. 2020. DOI: 10.5281/zenodo.3673058 (cit. on p. 56).

[219] L. Rettig, I. Vipavc, B. Tasic, K. Baumann, and S. Sigloch. *maDMPs in Linked Data Pipeline as part of Research Data Connectome*. Version 1.0.0. This report is a summary of work done by the Fancycatmeme team at the RDA Hackathon on machine- actionable Data Management Plans that took place between 27th and 29th May 2020. July 2020. DOI: 10.5281/zenodo.3944456 (cit. on p. 69).

[220] C. Ribeiro, J. R. da Silva, J. A. Castro, R. C. Amorim, and J. C. Lopes. " Projeto TAIL - Gestão de dados de investigação da produção ao depósito e à partilha (resultados preliminares)." In: *Cadernos BAD* N.2, jul-dez (2016), pp. 256–264. URL: https://www.bad.pt/publicacoes/index.php/cadernos/article/view/1603 (cit. on pp. 30, 31).

[221] C. Ribeiro, J. R. da Silva, J. A. Castro, R. C. Amorim, and J. C. Lopes. " Research Data Management Tools and Workflows: Experimental Work at the University of Porto." In: *IASSIST Quarterly* 42.2 (2018).

[222] J. Riley. "Understanding metadata." In: *Washington DC, United States: National Information Standards Organization* 23 (2017) (cit. on pp. 22, 24).

[223] K. Rimkus, T. Padilla, T. Popp, and G. Martin. "Digital preservation file format policies of ARL member libraries: an analysis." In: *D-Lib Magazine* 20.3/4 (2014) (cit. on p. 10).

[224] L. Rolando, J. Carlson, P. Hswe, S. W. Parham, B. Westra, and A. L. Whitmire. "Data management plans as a research tool." In: *Bulletin of the Association for Information Science and Technology* 41.5 (2015), pp. 43–45. DOI: 10.1002/bult.2015.1720410510 (cit. on p. 34).

[225] A. Sallans and M. Donnelly. "DMP Online and DMPTool: different strategies towards a shared goal." In: *International Journal of Digital Curation* 7.2 (2012). DOI: 10.2218/ijdc.v7i2.235 (cit. on pp. 53, 54).

[226] L. Salvatori, A. Petrus, N. Lambeng, and E. Blumer. "Keep calm and fill in your DMP: Lessons Learnt from a Swiss DMP-template initiative." In: *International Journal of Digital Curation* 13.1 (2018), pp. 215–222. DOI: 10.2218/ijdc.v13i1.617 (cit. on pp. 92, 94).

[227] S.-A. Sansone, P. McQuilton, P. Rocca-Serra, A. Gonzalez-Beltran, M. Izzo, A. L. Lister, and M. Thurston. "FAIRsharing as a community approach to standards, repositories and policies." In: *Nature biotechnology* 37.4 (2019), pp. 358–367 (cit. on p. 18).

[228] D. S. Sayogo and T. A. Pardo. "Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data." In: *Government Information Quarterly* 30 (2013), pp. 19–31. DOI: 10.1016/j.giq.2012.06.011 (cit. on pp. 1, 2).

[229] Q. Schiermeier. "For the record Making project data freely available is vital for open science." In: *Nature* 555.7696 (2018), pp. 403–405 (cit. on pp. 34, 36, 43, 45, 53).

[230] J. Schirrwagen, P. Cimiano, V. Ayer, C. Pietsch, C. Wiljes, J. Vompras, and D. Pieper. "Expanding the research data management service portfolio at bielefeld university according to the three-pillar principle towards data FAIRness." In: *Data Science Journal* 18.1 (2019). DOI: 10.5334/dsj-2019-006 (cit. on pp. 86, 94, 97).

[231] D. Schmitz, D. Hausen, and U. Trautwein-Bruns. *Content of a Data Management Plan. RWTH Aachen University*. DOI: 10.18154/RWTH-2019-10064 (cit. on p. 37).

[232] S. Scholtens, M. Jetten, J. Böhmer, C. Staiger, I. Slouwerhof, M. van der Geest, and C. W. van Gelder. *Final report: Towards FAIR data steward as profession for the lifesciences. Report of a ZonMw funded collaborative approach built on existing expertise*. Zenodo. 2019. DOI: 10.5281/zenodo.3474789.

[233] J. Schöpfel, C. Ferrant, F. André, and R. Fabre. "Research data management in the French national research center (CNRS)." In: *Data technologies and applications* (2018). DOI: 10.1108/DTA-01-2017-0005 (cit. on p. 102).

[234] C. Science, T. Board, E. National Academies of Sciences, Medicine, et al. "Active Data Management Plans as a Planning Tool." In: *Life-Cycle Decisions for Biomedical Data: The Challenge of Forecasting Costs*. National Academies Press (US), 2020. DOI: 10.17226/25639 (cit. on p. 49).

[235] SciLifeLab. *How to create a Data Management Plan using SciLifeLab DSW*. 2020. URL: https://www.youtube.com/watch?v=HY2DVnNGkAs&ab_channel=SciLifeLab (cit. on pp. 64, 65).

[236] K. Shearer and F. Furtado. *COAR Survey of Research Data Management: Results*. Confederation of OpenAccess Repositories, available at https://www.coar-repositories.org/files/COAR-RDM-Survey-Jan-2017.pdf. 2017 (cit. on p. 1).

[237] J. R. da Silva, J. A. Castro, C. Ribeiro, and J. C. Lopes. "Dendro: collaborative research data management built on linked open data." In: *European Semantic Web Conference*. Springer. 2014, pp. 483–487 (cit. on p. 26).

[238] J. R. da Silva, C. Ribeiro, and J. C. Lopes. "The Dendro Research Data Management Platform: Applying Ontologies to Long-Term Preservation in a Collaborative Environment." In: *Proceedings of the 11th International Conference on Digital Preservation, iPRES*. 2014 (cit. on pp. 3, 26).

[239] R. Rb-Silva and Y. Karimova. *aMILE: Application of text mining to clinical reports of patients with acute myeloid leukemia*. Jan. 2021. DOI: 10.5281/zenodo.4477657 (cit. on pp. 41, 112, 226).

[240] D. Simberloff, B. Barish, K. Droegemeier, D. Etter, N. Fedoroff, K. Ford, L. Lanzerotti, A. Leshner, J. Lubchenco, M. Rossmann, et al. "Long-lived digital data collections: enabling research and education in the 21st century." In: *National Science Board, National Science Foundation* (2005) (cit. on pp. 9, 34, 35).

[241] S. Simms and S. Jones. "Next-Generation Data Management Plans: Global, Machine-Actionable, FAIR." In: *International Journal of Digital Curation* 12.1 (2017). DOI: 10.2218/ijdc.v12i1.513 (cit. on pp. 2, 34, 37, 47, 53, 55, 65).

[242] S. Simms, S. Jones, D. Mietchen, and T. Miksa. "Machine-actionable data management plans (maDMPs)." In: *Research Ideas and Outcomes* 3 (2017), e13086. DOI: 10.3897/rio.3.e13086 (cit. on pp. 47–50, 52, 53).

[243] S. Simms, S. Jones, T. Miksa, D. Mietchen, N. Simons, and K. Unsworth. "A landscape survey of ActiveDMPs." In: *International Journal of Digital Curation* 13.1 (2018). DOI: 10.2218/ijdc.v13i1.629 (cit. on pp. 47, 55, 63, 65).

[244] S. Simms, M. Strong, S. Jones, and M. Ribeiro. "The Future of Data Management Planning: Tools, Policies, and Players." In: *International Journal of Digital Curation* 11.1 (2016), p. 208. ISSN: 1746-8256. DOI: 10.2218/ijdc.v11i1.413 (cit. on pp. 53–55).

[245] N. Smale, K. Unsworth, G. Denyer, and D. Barr. "The History, Advocacy and Efficacy of Data Management Plans." In: *bioRxiv* (2018). DOI: 10.1101/443499 (cit. on pp. 36, 44).

[246] C. Smith. "Controlled Vocabularies: Past, Present and Future of Subject Access." In: *Cataloging & Classification Quarterly* 59.2-3 (2021), pp. 186–202 (cit. on p. 24).

[247] C. Standards and C. Board. *CoreTrustSeal Requirements 2023-2025*. Version V01.00. Sept. 2022. DOI: 10.5281/zenodo.7051012 (cit. on p. 29).

[248] J. C. van Steen and O. Task Force. *Revised Field of Science and Technology (FOS) Classification in the Frascati manual. Working Party of National Experts on Science and Technology Indicators*. 2007. URL: https://www.oecd.org/sti/inno/38235147.pdf (cit. on p. 143).

[249] W. Stephanie. *Funder Data-Sharing Policies: Overview and Recommendations*. 2017. DOI: 10.6084/m9.figshare.5395456.v2 (cit. on pp. 16, 22).

[250] V. Stodden, V. Ferrini, M. Gabanyi, K. Lehnert, J. Morton, and H. Berman. "Open access to research artifacts: Implementing the next generation data management plan." In: *Proceedings of the Association for Information Science and Technology* 56.1 (2019), pp. 481–485 (cit. on pp. 35, 37, 71).

[251] M. Suchánek, R. Hooft, R. Pergl, K. Hettne, A. Terra, and M. Pejcoch. *Data Stewardship Wizard Import and Export and FAIR Implementation Profiles*. Version 1.0.1. This report is a summary of work done by the DSW team at the RDA Hackathon on machine-actionable Data Management Plans that took place between 27th and 29th May 2020. July 2020. DOI: 10.5281/zenodo.3944189 (cit. on pp. 65, 69).

[252] T. Sveinsdottir, J. Davidson, and V. Proudman. *An Analysis of Open Science Policies in Europe, v7*. Version 7. Apr. 2021. DOI: 10.5281/zenodo.4725817 (cit. on pp. 11–13).

[253] E. Svenonius. "Design of controlled vocabularies in the context of emerging technologies." In: *SRELS Journal of Information Management* 25.4 (1988), pp. 215–227 (cit. on p. 25).

[254] A. Swan and S. Brown. "To share or not to share: Publication and quality assurance of research data outputs." In: *Report commissioned by the Research Information Network (RIN). Annex detailed findings for the eight research areas (June 2008)* June (2008), p. 56 (cit. on p. 1).

[255] A. Tani, L. Candela, and D. Castelli. "Dealing with metadata quality: The legacy of digital library efforts." In: *Information Processing and Management* 49.6 (2013), pp. 1194–1205. DOI: 10.1016/j.ipm.2013.05.003 (cit. on p. 31).

[256] C. T. Team. "CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC." In: *Zenodo* (2020). DOI: 10.5281/zenodo.3820473 (cit. on pp. 44, 45).

[257] J. P. Tennant, F. Waldner, D. C. Jacques, P. Masuzzo, L. B. Collister, and C. H. Hartgerink. "The academic, economic and societal impacts of Open Access: an evidence-based review." In: *F1000Research* 5 (2016) (cit. on p. 12).

[258] C. Tenopir, E. D.Dalton, S. Allard, M. Frame, I. Pjesivac, B. Birch, D. Pollock, and K. Dorsett. "Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientistis Worldwide." In: *PLoS ONE* 10.8 (2015). DOI: 10.1371/journal.pone.0134826 (cit. on p. 1).

[259] C. Tenopir, S. Talja, W. Horstmann, E. Late, D. Hughes, D. Pollock, B. Schmidt, L. Baird, R. Sandusky, and S. Allard. "Research data services in European academic research libraries." In: *Liber Quarterly* 27.1 (2017). DOI: 10.18352/lq.10180 (cit. on pp. 102, 104).

[260] M. Töwe and C. Barillari. "Who does what?–Research data management at ETH Zurich." In: *Data Science Journal* 19.1 (2020), p. 36. DOI: 10.5334/dsj-2020-036 (cit. on pp. 85, 94).

[261] J. Trischler, S. J. Pervan, S. J. Kelly, and D. R. Scott. "The value of codesign: The effect of customer involvement in service design teams." In: *Journal of Service Research* 21.1 (2018), pp. 75–100. DOI: 10.1177/1094670517714060 (cit. on p. 4).

[262] D. Tziotzios, G. Kakaletris, G. Kalabokis, and E. Papadopoulou. *Import and Export of maDMPs from Argos/OpenDMP*. Version 1.0.0. This report is a summary of work done by the DMP Ninjas team at the RDA Hackathon on machine- actionable Data Management Plans that took place between 27th and 29th May 2020. July 2020. DOI: 10.5281/zenodo.3944314 (cit. on p. 69).

[263] UKRI. *Concordat on Open Research Data, UK Research and Innovation*. 2016 (cit. on p. 16).

[264] J. E. Van Loon, K. G. Akers, C. Hudson, and A. Sarkozy. "Quality evaluation of data management plans at a research university." In: *IFLA journal* 43.1 (2017), pp. 98–104. DOI: 10.1177/0340035216682041 (cit. on p. 45).

[265] M. Verburg, I. von Stein, L. Cepinskas, H. L'Hours, P. Herterich, J. Davidson, K. Ashley, O. Rouchon, A. Greco, S. M. Gallas, and S. P. Gaiarin. *D4.3 Report on the certification support and guidance for repositories and reviewers*. Version V1.0. July 2021. DOI: 10.5281/zenodo.6656437 (cit. on p. 29).

[266] T. H. Vines, A. Y. K. Albert, R. L. Andrew, F. Débarre, D. G. Bock, M. T. Franklin, K. J. Gilbert, J. S. Moore, S. Renaut, and D. J. Rennison. "The availability of research data declines rapidly with article age." In: *Current Biology* 24.1 (2014), pp. 94–97. DOI: 10.1016/j.cub.2013.11.014 (cit. on pp. 1, 15).

[267] C. H. Vitale and H. L. Moulaison Sandy. "Data Management Plans: A Review." In: *DESIDOC Journal of Library Information Technology* 39.6 (2019). DOI: 10.14429/djlit.39.06.15086 (cit. on pp. 2, 37, 47).

[268] V. Voß and G. Hamrin. "Quadcopters or Linguistic Corpora: Establishing RDM Services for Small-Scale Data Producers at Big Universities." In: *LIBER Quarterly: The Journal of the Association of European Research Libraries* 28.1 (Nov. 2018), pp. 1–58. DOI: 10.18352/lq.10255 (cit. on pp. 84, 99).

[269] M. Wali, N. Lvov, I. Hasani-Mavriqi, C. Ladurner, J. Vandendorpe, and N. Zimmer. *InvenioRDM and maDMPs integration*. Version 1.0.0. This report is a summary of work done by the DMP InvenioRDM team at the RDA Hackathon on machine- actionable Data Management Plans that took place between 27th and 29th May 2020. July 2020. DOI: 10.5281/zenodo.3944442 (cit. on p. 69).

[270] P. Walk, P. Neish, and T. Miksa. *WG DMP Common standards case statement*. 2017. URL: https://tinyurl.com/abk7nhjk (cit. on pp. 47, 48).

[271] R. E. Wallace. "What If You Don't Plan?" In: *Journal of Information Systems Management* 5.4 (1988), pp. 78–80. DOI: 10.1080/07399018808962947 (cit. on p. 36).

[272] S. L. Weibel. "The Dublin Core Metadata Element Set ISSN: 1041-5653." In: *World Spatial Metadata Standards*. Elsevier, 2005, pp. 493–513 (cit. on p. 22).

[273] H. C. White. "Descriptive Metadata for Scientific Data Repositories: A Comparison of Information Scientist and Scientist Organizing Behaviors." In: *Journal of Library Metadata* 14.1 (2014), pp. 24–51. DOI: 10.1080/19386389.2014.891896 (cit. on pp. 1, 30).

[274] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. a.C 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. a. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. "The FAIR Guiding Principles for scientific data management and stewardship." In: *Scientific Data* 3 (2016). DOI: 10.1038/sdata.2016.18 (cit. on pp. 1, 12, 17, 22, 23).

[275] T. Willaert, J. Cottyn, U. Kenens, T. Vandendriessche, D. Verbeke, and R. Wyns. "Research data management and the evolutions of scholarship: Policy, infrastructure and data literacy at KU Leuven." In: *LIBER Quarterly: The Journal of the Association of European Research Libraries* 29.1 (Sept. 2019), pp. 1–19. DOI: 10.18352/lq.10272 (cit. on pp. 83, 94, 97).

[276] C. Willis, J. Greenberg, and H. White. "Analysis and Synthesis of Metadata Goals for Scientific Data." In: *Journal of the Association for Information Science and Technology* 63.8 (2012), pp. 1505–1520. DOI: 10.1002/asi.22683 (cit. on pp. 9, 22–24, 39).

[277] C. Willoughby and J. G. Frey. "Data management matters." In: *Digital Discovery* 1 (3 2022), pp. 183–194. DOI: 10.1039/D1DD00046B (cit. on p. 14).

[278] J. Winn. "Open Data and the Academy: An Evaluation of CKAN for Research Data Management." In: *IASSIST 2013* May (2013), pp. 28–31. URL: http://eprints.lincoln.ac.uk/9778 (cit. on pp. 3, 25, 28, 31).

[279] J. Wittenberg and M. Elings. "Building a Research Data Management Service at the University of California, Berkeley: A tale of collaboration." In: *IFLA Journal* 43.1 (2017). DOI: 10.1177/0340035216686982 (cit. on pp. 2, 74).

[280] P. Wittenburg, H. Van de Sompel, J. Vigen, A. Bachem, L. Romary, M. Marinucci, T. Andersson, F. Genova, C. Best, W. Los, et al. *Riding the wave: How Europe can gain from the rising tide of scientific data*. 2010 (cit. on p. 11).

[281] I. Wolff, D. Broneske, and V. Köppen. "A first metadata schema for learning analytics research data management." In: *o-bib. Das offene Bibliotheksjournal/Herausgeber VDB* 8.4 (2021), pp. 1–20 (cit. on pp. 22, 23).

[282] *World Bank Group. Open Data for Sustainable Development*. Policy Note ICT01. 2015 (cit. on p. 1).

[283] *World Bank Group. World development report 2021: Data for better lives*. Policy Note ICT01. 2021 (cit. on pp. 8–10, 14, 15).

[284] M. Wrzeszcz, Ł. Dutka, R. G. Słota, and J. Kitowski. "New approach to global data access in computational infrastructures." In: *Future Generation Computer Systems* 125 (2021), pp. 575–589. DOI: https://doi.org/10.1016/j.future.2021.06.054 (cit. on p. 22).

[285] S. Xiao, T. Y. Ng, and T. T. Yang. "Research data stewardship at the University of Hong Kong." In: *Library Management* 43.1 (2021), pp. 128–147. DOI: 10.1108/LM-09-2021-0079 (cit. on pp. 82, 94, 97).

[286] N. Zimmer, N. Smith, Y. Ebrahim, P. Neish, J. Philipson, B. Riley, and L. Winton. *Integrating a Converis CRIS/RIMS with DMPRoadmap and Export/Import maDMP from Figshare*. Version 1.0.0. This report is a summary of work done by the maDMP Link team at the RDA Hackathon on machine- actionable Data Management Plans that took place between 27th and 29th May 2020. July 2020. DOI: 10.5281/zenodo.3944381 (cit. on p. 69).

[287] J. J. Zondergeld, R. H. Scholten, B. M. Vreede, R. S. Hessels, A. Pijl, J. E. Buizer-Voskamp, M. Rasch, O. A. Lange, and C. L. Veldkamp. "FAIR, safe and high-quality data: The data infrastructure and accessibility of the YOUth cohort study." In: *Developmental Cognitive Neuroscience* 45 (2020), p. 100834. DOI: https://doi.org/10.1016/j.dcn.2020.100834 (cit. on pp. 80, 97).

[288] M. Zozus. *The data book: collection and management of research data*. Chapman and Hall, 2017 (cit. on pp. 9, 10, 15, 17, 21, 22).

A APPENDIX

A.1 CONTRIBUTIONS

A.1.1 DMPs

1. DMP Gamma radiation monitoring campaign at the Azores ENA-ARM station (Graciosa Island)[1], 2019;
2. DMP FARSYD - FArming SYstems as tool to support policies for effective conservation and management of high nature value farmlanDs[2], 2019;
3. DMP Sail, 2021 [15];
4. DMP Facilitated Genome Editing as Responsible Research and Innovation, 2021 [212];
5. DMP SCReLProg (Self and Co-regulation in e-Learning of Computer Programming), 2021 [143];
6. DMP aMILE: Application of text mining to clinical reports of patients with acute myeloid leukemia, 2021 [239];
7. DMP VigilHate - VIGILANT CITIZENS AGAINST HATE: How to counter bystander apathy and increase citizens' commitment against online hate speech?, 2021 [140];
8. DMP Neurophysiological bases of decision-making processes: dissociating risk and uncertainty in the human brain, 2022 [139];
9. DMP Life Skills VR: Life Skills for Employment in COVID-19 Era through VR Innovation, 2022 [142];
10. DMP FronTowns: Think big on small frontier towns: Alto Alentejo and Alta Extremadura leonesa (13th - 16th centuries), 2022 [147];

A.1.2 Posters

1. Poster "Controlled vocabularies in the description of research data on Dendro", Information Science Days, FEUP, Porto, 2016;
2. Poster "Método colaborativo entre curadores e investigadores na preparação de um Plano de Gestão de Dados", CONFOA Conference 9, Portugal, Lisboa, 2018;
3. Poster "The collaborative method between curators and researchers in the preparation of a Data Management Plan and Privacy Impact Assessment"[3], 14th RDA Plenary, Finland, Helsinki, 2019;
4. Co-author in poster "Narratives on the adoption of research data management practices in Portuguese research groups", 16th RDA Plenary, Costa Rica, 2020;
5. Poster "Research Data Management Workflows: Steps towards Implementing maDMPs", 16th RDA Plenary, Costa Rica, 2020;
6. Poster "Controlled Vocabularies and Domain Data Protocols for the Social Sciences: the application of the collaborative building method". 19th RDA Plenary, Seoul, 2022;

1 https://tinyurl.com/vusz7yca
2 https://tinyurl.com/2f4hw6kw
3 https://www.rd-alliance.org/system/files/dmp_pia_method_yuliakarimova_41_.png

A.1.3 Presentations

1. Presentation "Controlled vocabularies in the description of research data in Dendro", 7th Luso-Brazilian Conferência Open Access (CONFOA), Portugal, Viseu, 2016;

2. Presentation "Ferramentas de descrição e anotação no workflow GDI: caso do estudo com o Dendro e serviços EUDAT"[4], 3º Forum GDI, 2017;

3. Presentation "Flexible metadata models and controlled vocabularies for research data description in multiple domains", TPDL and Dublin Core Conference, Portugal, Porto, 2018 [132];

4. Presentation "Using B2NOTE: the U.Porto pilot", EUDAT Conference "Putting the EOSC vision into practice", Portugal, Porto, 2018;

5. Presentation "O método colaborativo entre curadores e investigadores na preparação de um DMP e PIA", 5º Forum Gestão de dados de investigação, Portugal, Aveiro, 2019;

A.1.4 Publications

1. Paper Karimova, Y., Castro, J. A. Vocabulários controlados na descrição de dados de investigação no Dendro. Cadernos BAD, (2), 241-255, 2016 [134];

2. MSc dissertation "Vocabulários controlados na descrição de dados de investigação no Dendro" (Controlled vocabularies for describing research data in Dendro), 2016 [133];

3. Co-author of the Book chapter: Castro, J. A., Amorim, R. C., Karimova, Y., da Silva, J. R., Gattelli, R., Ribeiro, C. Involving Data Creators in an Ontology-Based Design Process for Metadata Models. In Developing Metadata Application Profiles, pp. 181-214, IGI Global, 2017 [38];

4. "Best Student Paper Award" in Proceedings of Scientific Meetings: Karimova, Y., Castro J.A., da Silva, J.R., Pereira N., Ribeiro C. Promoting semantic annotation of research data by their creators: A use case with B2NOTE at the end of the RDM workflows. In: Garoufallou E., Virkus S., Siatri R., Koutsomiha D. (eds) Metadata and Semantics Research, 11th International Conference, MTSR 2017. Communications in Computer and Information Science, vol 755. Springer, Cham, 2017 [136];

5. Paper in International Journal: Karimova, Y., Castro J.A., da Silva, J.R., Pereira N., Rodrigues, J. and Ribeiro C. Description + Annotation: semantic data publication workflow with Dendro and B2NOTE. International Journal of Metadata, Semantics and Ontologies, Vol.12, N.4. DOI:10.1504/IJMSO.2017.093645, 2018 [137];

6. Co-author of the paper: Rodrigues J., Castro, J.A., Silva, J.R., Pereira, N., Karimova, Y. and Ribeiro, C. Domain-specific metadata for publishing research data in B2SHARE. Cadernos BAD, n.13, 2018;

7. Flexible metadata models and controlled vocabularies for research data description in multiple domains. In: Bulletin of IEEE Technical Committee on Digital Libraries, Volume 15, Issue 2, February 2019;

8. Paper Karimova Y., Castro J.A., Ribeiro C. Data Deposit in a CKAN Repository: A Dublin Core-Based Simplified Workflow. In: Manghi P., Candela L., Silvello G. (eds) Digital Libraries: Supporting Open Science, IRCDL 2019. Communications in Computer and Information Science, vol 988. Springer, 2019 [135];

9. Report of the participation in the RDA Hackathon on maDMPs[5] "Research Data Management Workflows and maDMPs", 2020 [141];

10. Paper Karimova Y., Ribeiro C., David G. Institutional Support for Data Management Plans: Five Case Studies. In: Garoufallou E., Ovalle-Perandones MA. (eds) Metadata and Semantic Research. MTSR 2020. Communications in Computer and Information Science, vol 1355. Springer, Cham, 2021 [146];

4 http://forumgdi.rcaap.pt/wp-content/uploads/2017/07/08-3ForumGDI_Ferramentas_descricao.pdf
5 https://www.rd-alliance.org/rda-hackathon-madmps

11. Paper Karimova, Yulia, Cristina Ribeiro, and Gabriel David. "Institutional support for data management plans: case studies for a systematic approach." International Journal of Metadata, Semantics and Ontologies, 15.3, p.178-191, 2021 [145];

12. Co-author of the Book: Engelhardt, Claudia, Biernacka, Katarzyna, Coffey, Aoife, Cornet, Ronald, Danciu, Alina, Demchenko, Yuri, Downes, Stephen, Erdmann, Christopher, Garbuglia, Federica, Germer, Kerstin, Helbig, Kerstin, Hellström, Margareta, Hettne, Kristina, Hibbert, Dawn, Jetten, Mijke, Karimova, Yulia, Kryger Hansen, Karsten, Kuusniemi, Mari Elisa, Letizia, Viviana, ... Zhou, Biru. D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions (V1.2.1) [Computer software]. Zenodo, 2022 [80];

A.1.5 Workshops

1. Speaker on the Workshop "Hands-on DMP. Bio Data Management"[6], Lisbon, 2019;

2. Speaker on the Workshop "Introduction to Research Data Management. Training School of the European Network on Individualized Psychotherapy Treatment of Young People with Mental Disorders (TREATme)", Krakow, 2019 [21];

3. Co-organization and speaker on the Workshop "Open Science Workshop"[7], i3S, Porto, 2021;

4. Speaker on the Webinar Fair Data: desafios e oportunidades, i3S[8], 2021;

A.1.6 Other contributions

1. Development of the RDM@Uporto website[9] with a series of trainings, guidelines, and other useful materials related to RDM to help researchers, 2016;

2. Participation in the FAIRsFAIRs Calls for "Support towards achieving CoreTrustSeal certification"[10] and "Support towards increasing repository interoperability"[11], 2019;

3. Collaboration with the Portuguese Institute of Oncology of Porto (IPO-Porto) on Biaward call https://biaward.pt/, 2021;

4. Enrollment in the group Open Science trainers[12], 2021;

5. Collaboration with the INESC TEC Brussels HUB[13] and development of the training materials related to the RDM activities, including DMP creation, 2021;

6. Elaboration the MOOC about DMP "Planos de Gestão de Dados (DMP): desmistificação" with the University of Évora (ongoing).

6 https://biodata.pt/node/100
7 https://www.i3s.up.pt/training-detail.php?v=160
8 https://sigarra.up.pt/feup/pt/noticias_geral.ver_noticia?p_nr=118950
9 https://rdm.up.pt/
10 https://www.fairsfair.eu/form/support-towards-achieving-coretrustseal-certification
11 https://www.fairsfair.eu/form/support-towards-increasing-repository-interoperability
12 https://www.fosteropenscience.eu/trainers-directory
13 https://hub.inesc.pt/

B APPENDIX

B.1 INTERVIEW PROTOCOL FOR DMP SUPPORT MEETINGS (1ST VERSION)

1st meeting (contextualization with project, general questions):

Organizational questions

1. Presentation of my thesis, goals and brief description
2. Type of support that data steward can offer (DMP creation, DMP monitoring or analysis of the existing DMP for improvements or publishing)
3. Confidentiality agreement

Experience in RDM and DMP

1. Experiences in DMP creation or publishing
2. Experiences in data deposit, description, sharing, reusing, or other RDM activities
3. Assisted workshops or training related to the RDM and DMP

Information about project

1. Title
2. Begin and end date of the project
3. Budget
4. Brief description
5. Scientific domain
6. Countries (partners) of the project
7. Data that will be collected, created or reused on the project
8. Methodology and tools for data collection

Additional questions

1. Contact with the DPO
2. Contact with Ethics Committee
3. Documents such as informed consent, agreements, etc.

Next meetings after the data steward analyzed all documents (more detailed specific questions):

Questions related to the datasets

1. Data organization
2. Restrictions, limitations, licenses
3. Data repository
4. Data storage

Questions related to the project members

1. Responsibilities
2. Access rights
3. Data sharing between partners and additional documents needed for this purpose, if necessary
4. Owner issues

Additional questions

1. Specific terms used on the project, for example, to describe tools, methodologies, and characteristics, among others, that can be used as controlled vocabularies and Domain Data Protocols for DMP.

B.2 INTERVIEW PROTOCOL FOR DMP SUPPORT MEETINGS (IMPROVED VERSION)

1st meeting (importance, motivation, contextualization with project, general questions):

Motivation and importance

1. Motivating researchers to create a plan
2. Emphasizing that DMP creation/monitoring processes take time and effort on both sides, as the data steward cannot create a plan without the participation of the researchers
3. Emphasizing the fact that in some cases, an authorization may be required from all project participants, both to create a plan and to publish it
4. Verification of whether there is any version of the plan
5. Clarification of the issues about the necessity and benefits of the DMP creation/monitoring/publication

Organizational questions

1. Presentation of my thesis, goals and brief description
2. Type of support that data steward can offer (DMP creation, DMP monitoring or analysis of the existing DMP for improvements or publishing, DMP-light version creation)
3. Providing examples of already created DMPs, preferably for the same scientific domain
4. Confidentiality agreement

Experience in RDM and DMP

1. Experiences in DMP creation or publishing
2. Experiences in data deposit, description, sharing, reusing, or other RDM activities
3. Assisted workshops or training related to the RDM and DMP

Information about project

1. Title
2. Begin and end date of the project
3. Budget
4. Brief description
5. Scientific domain

6. Countries (partners) of the project
7. Data that will be collected, created or reused on the project
8. Methodology and tools for data collection

Additional questions

1. Contact with the DPO
2. Contact with Ethics Committee
3. Documents such as informed consent, agreements, etc.

Next meetings after the data steward analyzed all documents (more detailed specific questions):

Questions related to the datasets

1. Data organization
2. Restrictions, limitations, licenses
3. Data repository
4. Data storage

Questions related to the project members

1. Responsibilities
2. Access rights
3. Data sharing between partners and additional documents needed for this purpose, if necessary
4. Owner issues

Additional questions

1. Specific terms used on the project, for example, to describe tools, methodologies, and characteristics, among others, that can be used as controlled vocabularies and Domain Data Protocols for DMP.

C APPENDIX

C.1 EMAIL PARA CONTACTO (PT)

Boa tarde,

Devido aos novos requisitos de financiadores, os DMP (Planos de Gestão de Dados) tornam-se tarefas que o investigador tem de incluir na proposta ou realizar no decorrer do projeto. São tarefas que requerem algum esforço e tempo. Por isso, venho por este meio perguntar se está interessado em receber o nosso apoio na criação de DMP para o seu projeto que foi aprovado em X, no acompanhamento (monitorização) do plano e no esclarecimento das dúvidas ligadas à gestão de dados de investigação.

Além disso, podemos ajudar no depósito e partilha de dados do projeto, com intuito de tornar os seus dados pesquisáveis, acessíveis, interoperáveis e reutilizáveis (FAIR), aumentando assim a visibilidade e o valor do seu trabalho científico, usando o repositório do INESC TEC (rdm.inesctec.pt) com atribuição de DOI.

Com os melhores cumprimentos,

Yulia Karimova, Investigadora e Data Steward de INESC TEC, FEUP.

C.2 EMAIL FOR CONTACT (EN)

Good afternoon,

Due to new funder requirements, DMP (Data Management Plans) become tasks that the researcher has to include in the proposal or perform during the course of the project. These are tasks that require some effort and time. Therefore, I hereby ask if you are interested in receiving our support in the creation of a DMP for your project that was approved in X, in the follow-up (monitoring) of the plan and in the clarification of issues related to the management of research data.

In addition, we can help with the deposit and sharing of project data, in order to make your data findable, accessible, interoperable and reusable (FAIR), thus increasing the visibility and value of your scientific work, using the INESC TEC repository (rdm.inesctec.pt) with a DOI assignment.

Best regards,

Yulia Karimova, Researcher and Data Steward at INESC TEC, FEUP.

D APPENDIX

D.1 ACORDO DE CONFIDENCIALIDADE (PT)

Acordo de Confidencialidade no Âmbito da Tese de Doutoramento "Research data description in multiple domains: supporting researchers with data management plans" (PT)

Entre

YULIA KARIMOVA - Estudante de doutoramento da FEUP e Data Steward de INESC TEC

E

(responsável do projeto por parte de INESC TEC, título de projeto)

O presente acordo tem por objecto garantir a confidencialidade e proteção da informação classificada como protegida, confidencial ou outra de igual significado, trocada entre as partes com a seguinte finalidade: Discussões de interesse mútuo sobre o desenvolvimento de um Plano de Gestão de Dados (Data Management Plan (DMP)) para projeto acima referido, com a necessidade, neste contexto, de troca de informações entre as partes, que assumem natureza reservada.

A informação protegida, confidencial ou outra de igual significado não será usada, divulgada ou partilhada a qualquer título e de qualquer forma, em Portugal ou no estrangeiro.

A informação pública como: Título, Orçamento, Financiador, Entidade responsável do projeto, Breve descrição do contexto e objetivos, Domínio científico, Datas de início e de final, Nome da pessoa responsável, será publicada na tese de doutoramento.

A informação recolhida relacionada com o processo de criação e de acompanhamento de DMP como: experiência em gestão de dados, dificuldades durante a criação, pontos específicos dependentes de domínio, breve descrição de criação de DMP e o acompanhamento do mesmo, estado do DMP criado (no caso de o DMP ser publicado, o link/DOI de DMP) será publicada após a sua autorização.

As partes reconhecem e aceitam as disposições do presente Acordo.

Data

D.2 CONFIDENTIALITY AGREEMENT (EN)

Confidentiality Agreement under the Doctoral Thesis "Research data description in multiple domains: supporting researchers with data management plans"

Between

YULIA KARIMOVA - PhD student at FEUP and Data Steward at INESC TEC

And

(Name of responsible of the project on the part of INESC TEC, project title)

The purpose of this agreement is to guarantee the confidentiality and protection of information classified as protected, confidential or other of equal meaning, exchanged between the parties for the following purpose: Discussions of mutual interest on the development of a Data Management Plan (DMP) for the aforementioned project, with the need, in this context, of information exchange between the parties, which assume a reserved nature.

Protected, confidential or other information of equal meaning will not be used, disclosed or shared with anyone and in any way, in Portugal or abroad. Public information, such as: Title, Budget, Funder, Entity in charge of the project, Brief description of the context and objectives, Scientific domain, Start and end dates, Name of the person in charge will be published in the doctoral thesis.

The information collected related to the DMP creation and monitoring process, such as: experience in data management, difficulties during creation, specific domain-dependent aspects, brief description of DMP creation and its monitoring, DMP status (in case a DMP is published, the DMP link/DOI) will be published after your authorization.

The parties acknowledge and accept the provisions of this Agreement.

Date

E APPENDIX

E.1 CLASSIFICATION OF PROJECTS IN THE SAMPLE ACCORDING TO FOS

Field	Sub-Field	Preliminary	Actual
1. Natural sciences	1.1. Mathematics		2
	1.2. Computer and information sciences	22	15
	1.3. Physical sciences	2	2
	1.4. Chemical sciences		
	1.5. Earth and related environmental sciences	1	2
	1.6. Biological sciences	3	3
	1.7. Other natural sciences	2	2
2.Engineering and technology	2.1. Civil engineering		
	2.2. Electrical engineering, Electronic engineering, Information engineering	6	8
	2.3. Mechanical engineering	1	1
	2.4. Chemical engineering		
	2.5. Materials engineering		
	2.6. Medical engineering		
	2.7. Environmental engineering	6	5
	2.8. Environmental biotechnology	1	1
	2.9. Industrial biotechnology		
	2.10. Nano-technology		
	2.11. Other engineering and technologies	10	9
3.Medical and Health sciences	3.1. Basic medicine		1
	3.2. Clinical medicine		
	3.3. Health sciences	4	3
	3.4. Medical biotechnology	2	1
	3.5. Other medical sciences	1	
4.Agricultural sciences	4.1. Agriculture, Forestry, and Fisheries		3
	4.2. Animal and Dairy science		
	4.3. Veterinary science		
	4.4. Agricultural biotechnology		
	4.5. Other agricultural sciences	3	
5.Social sciences	5.1. Psychology		
	5.2. Economics and Business	8	8
	5.3. Educational sciences	1	
	5.4. Sociology	1	4
	5.5. Law		
	5.6. Political science	2	5
	5.7. Social and economic geography	1	3
	5.8. Media and communications		
	5.9. Other social sciences	1	
6.Humanities	6.1. History and Archaeology	1	1
	6.2. Languages and Literature		
	6.3. Philosophy, Ethics and Religion		
	6.4. Arts (arts, history of arts, performing arts, music)		
	6.5. Other humanities		

F APPENDIX

F.1 INSTITUTIONAL RESEARCH DATA MANAGEMENT POLICY

The Research Data Management (RDM) Policy is a central aspect of good scientific practice at the institution and focuses on ensuring that research data collected/produced during the projects are in compliance with RDM and funders' requirements such as openness and FAIRness. RDM constitutes the entire process throughout the lifecycle of data, from their collection/creation to their re-use and preservation. Moreover, all aspects related to the data need to be analyzed considering different factors like ethical and legal issues, and intellectual property rights. According to the H2020 Program Guidelines on FAIR Data, data should be *"as open as possible and as closed as necessary"*. The key points of this RDM Policy are:

1. data collected/produced during the projects are recognized as an important research output;

2. all researchers must be aware of the RDM and funder requirements, and make their research data according to FAIR principles;

3. the institution must provide all necessary support and guidance to help researchers in RDM issues;

4. the institution must promote awareness of the value and importance of RDM among researchers and staff, and contributes to the organization of the necessary training and materials to improve knowledge;

5. the institution must analyze researchers difficulties and needs regarding to RDM issues;

6. the institution must provide appropriate tools and implement best practices to facilitate RDM on the projects; encourage researchers for good data management; meet the requirements of funders and publishers for research data management and sharing;

7. a Data Management Plan (DMP) must be created for each research project. Its creation is supported by the DMP Support System implemented in the institution;

8. the Principal Investigator must collaborate with the Data Steward to make the project in compliance with the RDM requirements and project results in accordance with the FAIR principles;

9. researchers can request support to clarify issues related to RDM issues from the DPO team, IT department, and Data Steward at their institution;

10. the DMP is a living document that must be monitored and improved throughout the project;

11. data collected/produced during the project can be deposited in an appropriate data repository for a long-term management and preservation. A repository from a list of trustworthy repositories[1] that can assign a DOI to each deposited dataset is preferred;

12. the institutional RDM policy must be publicly available on the institutional website and distributed among researchers in all clusters/departments;

13. the Data Steward reports periodically to the administration of the institution;

1 https://www.coretrustseal.org/why-certification/certified-repositories/

14. if in doubt about RDM issues or to request support, contact Data Steward by email: *xx@xx.com*, or phone number: *XXX-XX-XXX*, or personally on office Nº: *XXXX*.

This RDM policy will be regularly reviewed and improved in accordance with changes related to the RDM that occurred in the scientific community.

G APPENDIX

G.1 SURVEY QUESTIONS WITH THE FULL ANSWERS

Question (Figure 90): Based on our collaboration, did you (intend to) publish, present, or create any paper,presentation, poster, or workshop related to research data and/or DMP creation issues? Ifyes, please select all that apply:

- Paper;
- Poster;
- Presentation;
- Workshop;
- Course for others;
- Support for others in RDM and DMP issues;
- None;
- I am interested in publishing in the various types described. But in a co-author position. In other words, I am available to write the articles but I need the support. The lead of the article would have to be the person who accompanies us. Well, the focus of my research is not on DPM. Therefore, given the time available, we have to prioritize. But always available to collaborate and co-author;
- report or as article elements;
- We are publishing one on anonymization procedures, not DMP;
- not yet but it's a possibility.

Question (Figure 91): What did the data steward do for your project concretely? "Data Steward helped me to...":

- define preservation/sharing/accessibility issues;
- clarify doubts related to sensitive/personal/private data;
- define data repository for data publishing;
- create additional documentation, like DPIA, agreements between partners, authorization by Ethics Committee, informed consent, etc.;
- avoid undesirable unforeseen events and predict possible problems in the initial stages of a project;
- engage you in RDM activities and acquire RDM knowledge;
- clarify doubts related to the owner authorship issues;
- clarify doubts related to the ethical, legal, copyright, intellectual property rights, and licenses issues;
- make project data FAIR (Findable, Accessible, Interoperable, and Reusable);
- make the DMP of higher quality and more detailed;
- diminish time and effort for the DMP creation;
- improve my knowledge about managing my data on the projects;
-do my project in compliance with RDM and funder requirements;
- analyze different risks related to the research data;
- find more efficient ways of resources' usage and more adequate tools;
- raise the visibility of the projects and research results (increasing citation);
- so far there was not much interaction given that this is a recent service. Hopefully this will change in the next versions and projects.

Question (Figure 102): Please indicate your RDM needs at the institution (select all that apply):

- Support researchers in RDM;
- Detailed documentation / guidance / practicies / examples;
- Tools for RDM;
- Workshops / training sessions;
- Well-planned course considering the main recommendations of funders;
- Note: I had support but not from my institution. I had support from the INESC TEC scholarship.