

# Fostering the Adoption of DMP in Small Research Projects through a Collaborative Approach: a Case Study

André Maciel  
Faculty of Engineering, University of Porto

João Aguiar Castro  
INESC TEC, Porto, Portugal / Faculty of  
Engineering, University of Porto

Cristina Ribeiro  
INESC TEC, Porto, Portugal / Faculty of  
Engineering, University of Porto

Marta Almada  
UCIBIO/REQUIMTE,  
PORTO4AGEING – Competences Centre  
on Active Healthy Ageing of the University  
of Porto / Faculty of Pharmacy of the  
University of Porto, Portugal

Luís Midão

UCIBIO/REQUIMTE, PORTO4AGEING – Competences Centre on Active Healthy  
Ageing of the University of Porto / Faculty of Pharmacy of the University of Porto, Portugal  
/ Institute of Biomedical Sciences Abel Salazar, University of Porto, Portugal

## Abstract

To promote sound management of research data the European Commission, under the Horizon 2020 framework program, is promoting the adoption of a Data Management Plan (DMP) in research projects. Despite the value of a DMP to make data findable, accessible, interoperable and reusable (FAIR) through time, the development and implementation of DMPs is not yet a common practice in health research. Raising the awareness of researchers in small projects to the benefits of early adoption of a DMP is, therefore, a motivator for others to follow suit. In this paper we describe an approach to engage researchers in the writing of a DMP, in an ongoing project, FrailSurvey, in which researchers are collecting data through a mobile application for self-assessment of fragility. The case study is supported by interviews, a metadata creation session, as well as the validation of recommendations by researchers. With the outline of our process we also outline tools and services that supported the development of the DMP in this small project, particularly since there were no institutional services available to researchers.

*Submitted 18 July 2022 ~ Accepted 18 July 2022*

Correspondence should be addressed to João Aguiar Castro. Email: [joao.a.castro@inesctec.pt](mailto:joao.a.castro@inesctec.pt)

This paper was presented at International Digital Curation Conference IDCC22, online, 13-16 June, 2022

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution License, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



## Introduction

Through the introduction of increasingly sophisticated tools in the research process the complexity of data produced is growing. Hence, research environments have to keep up with the fast pace at which data are generated (Hey et al., 2009), resulting in a decline of data availability after publications (Vines et al., 2014). Lack of awareness regarding data sharing and publication opportunities may also hinder the availability of data over time.

In this context, the European Commission (EU), are asking grant applicants to write a DMP as a requirement for funding (European Commission, 2016). To comply with the DMP, researchers have to declare, among other things, how data will be handled during the research project, how it will be documented and made available. This effort aims to maximize data findability, accessibility, and interoperability and to improve the reusability of data generated by Horizon 2020 projects. This is the case of FAIR4Health,<sup>1</sup> which aims to promote the reuse and sharing of health data. It is therefore necessary to raise general awareness of the benefits of developing a DMP, since it is a proper channel for ensuring compliance with policies that encourage open access to research data (Simms and Jones, 2017). There are various tools that support researchers in the development of their DMP. These include DMPonline,<sup>2</sup> which provides a template with tailored guidance and example answers, with respect to funding agencies requirements; The Data Asset Framework (DAF),<sup>3</sup> a methodology with four major steps to improve the effectiveness of organizations in data management; and DMPTool,<sup>4</sup> an open-source, online application that enables researchers to create their plan with according to funding requisites and presents best practices recommendations.

Despite the progress made in recent years, the adoption of RDM best practices in small projects is not yet systematic. A study of DMPs effectiveness in Australian universities (Smale et al., 2018) concluded that there is no evidence that the development of a DMP brings professional benefits to researchers, since 64 percent of the DMPs had incomplete sentences, 55 percent lacked clarity in the type of data and 63 percent were not searchable by third parties. Moreover, Bishoff and Johnston (2015) concluded that researchers' data sharing strategies are inconsistent, and that more education is needed to ensure that they clearly implement data sharing actions in their DMP. On the other hand, Green, Cairns and White (2019), found that the main reason for researchers to develop their DMP is the existence of university policies and procedures that require it.

Therefore, the development of use cases to further engage communities in RDM is part of the recommendations to turn FAIR into reality (European Commission, 2018).

In this work we describe the steps to engage researchers from a small project in data management, through a set of collaborative activities between data curators and researchers. This process led to the design of a DMP for the FrailSurvey project. The goal of the project is the validation of a mobile application for self-assessment of frailty based on the Groningen Frailty Indicator,<sup>5</sup> in the Portuguese population.

In this paper we outline the proposal of a collaborative approach to develop DMPs for small projects, followed by a detailed description of the different approach steps, applied to the FrailSurvey project. During this case study, FrailSurvey's researchers, Marta Almada and Luís Midão, were invited to participate in the writing of this paper as a way to further improve their awareness and strengthen the collaboration between different RDM stakeholders.

---

<sup>1</sup> <https://www.fair4health.eu/>

<sup>2</sup> <https://dmponline.dcc.ac.uk/>

<sup>3</sup> <https://www.data-audit.eu/index.html>

<sup>4</sup> [https://dmptool.org/about\\_us](https://dmptool.org/about_us)

<sup>5</sup> [https://ec.europa.eu/eip/ageing/repository/frailsurvey-mobile-phone-app-self-assessment-frailty-status-among-community-dwelling\\_en](https://ec.europa.eu/eip/ageing/repository/frailsurvey-mobile-phone-app-self-assessment-frailty-status-among-community-dwelling_en)

## RDM Attitudes and Practices

In recent years, data sharing has grown and the same has happened to the number of RDM guidelines. Research organizations have progressively implemented strategies aimed at Open Science, emphasizing the idea of open access to research data and not just for publications (Zenk-Möltgen et al., 2018).

The adoption of suitable RDM practices is, on the one hand, closely linked to the willingness of researchers to share data, while on the other hand, there still seems to be a lack of knowledge among researchers regarding the services and tools they have at their disposal to improve their practices. Tenopir et al. (2011) ran an international survey, with a total of 1329 researchers, and concluded that data sharing was hampered by insufficient time and lack of funding. Curiously, most respondents (85 percent) were interested in using other researchers' data if the data were easily accessible, but only half reported making their data available. A more recent survey (Tenopir et al., 2015) has revealed a more positive perception and progress in data sharing behaviours, as most researchers reported making at least part of their data available to others. Arzberger et al. (2004) also verified that the lack of time and institutional support for data management were among the main reasons for researchers to retain data.

Another aspect to limit data availability to others is that peer scrutiny may expose errors or produce conclusions that contradict the original authors. According to Wicherts et al. (2011) the intention to share data can have an influence on how researchers manage their data, since those who apply greater diligence in the archiving and management of their data tend to commit fewer mistakes. Thus, starting a project with the intention to make the data eventually available to third parties can lead to the production of better-quality data, in line with FAIR principles. An institutional study with researchers from different disciplines have found that researchers from the basic sciences were the most familiar with funding agency requirements for DMPs, while, at the same time, they are also the ones most likely to share data outside their groups and publish in data repositories (Akers and Doty, 2013). However, a 2016 study, with 1317 researchers, concluded that there was no positive correlation between the funding agencies' policies for data sharing and data sharing attitudes (Kim and Stanton, 2016).

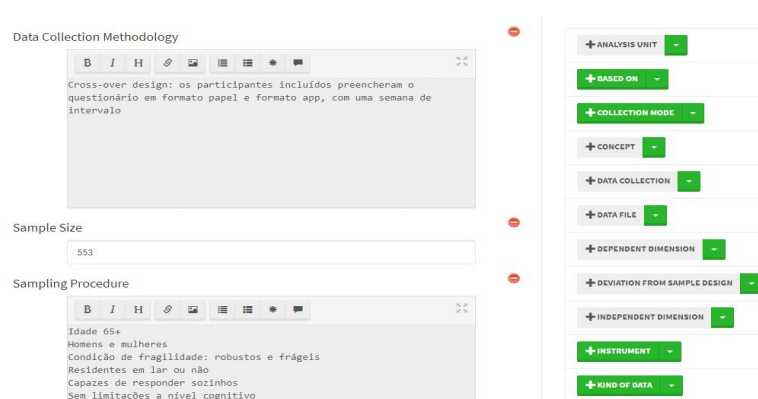
Wiley and Kerby (2018) carried out an institutional study, with graduate students and postdoctoral researchers, to evaluate RDM skills, and concluded that many researchers expressed frustration when former colleagues leave without providing annotations of the completed work. Consistent data description and organization was regarded as a challenge given the different workflows, practices and value concepts of individuals. A practical solution to address this challenge was the provision of short descriptions to enable group members to understand the research workflow. In another study, which consisted of 13 interviews with social scientists to assess factors of influence on researchers' perceptions and experiences in attempts to reuse data, it was concluded that data documentation was, among others, an important enabling factor for data reuse (Curty, 2016).

From interviews, carried out in 2016, with 23 quantitative social science researchers who have failed data reuse experiences (Yoon, 2016), it was found that access and interoperability are chief primary conditions for a successful data reuse experience. Although data documentation was less of an issue, at least for experienced researchers, the process was still seen as challenging. The lack of support was the most prominent issue of reported failed data reuse experiences, making it necessary to establish support systems for those willing to reuse data.

## Approach to the Design of a DMP for Small Research Projects

The approach we propose for the development of a DMP encompasses various collaborative activities between data curators and research groups, which include understanding the research processes, and the proposal of recommendations and their validation by the researchers. The steps were carried out with two members from the FrailSurvey. The number of researchers is justified by the size of the project and the inability to meet with other members. However, these two participants are very active in data collection and are those responsible for the management of the project data.

1. The process started with a study and familiarization with the scientific area of the FrailSurvey project. Therefore, the first step was the development of a script to support the interview with the researchers. The questions that constituted the script were based on RDM guidelines and tools. More specifically, the Curation Lifecycle Model of the Digital Curation Centre, provided the framework to structure the interview accordingly, complemented with the semi-structured Data Curation Profile Toolkit, Interview Sheet (Carlson, 2010) which provides questions for an RDM diagnostic, as well as others about researchers practices and perspectives.
2. The second moment in the interaction with researchers is training in metadata production. In this case, the FrailSurvey researchers described one of their datasets in a collaborative RDM platform for small research groups, developed at the University of Porto, Dendro (Rocha da Silva, 2016), as shown in Figure 1. Dendro was designed to support data description from the moment that data is created and uses Linked Open Data at the core. Its data model encourages data curators to model ontologies that can satisfy the description needs of each specific domain while retaining interoperability characteristics of the ontology itself. Considering the nature of the data produced in the FrailSurvey, the Data Documentation Vocabulary<sup>6</sup> after consultation in the Metadata Standards Directory,<sup>7</sup> was considered the most suitable for this project and was, therefore, recommended to researchers at the beginning of the data description session.



**Figure 1.** Example of data description by FrailSurvey researchers in the Dendro platform. On the right: metadata elements from the Data Documentation Initiative. On the left: Some metadata field completed by the researchers.

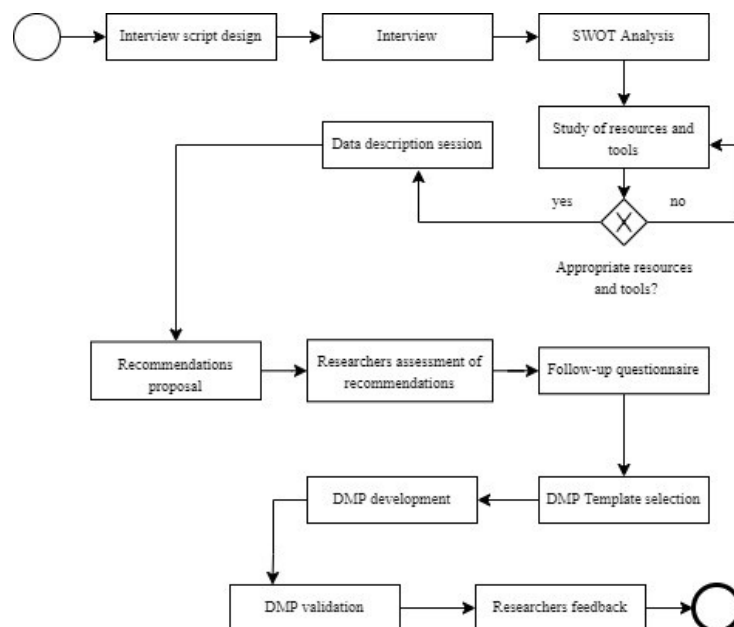
<sup>6</sup> <http://www.ddialliance.org/>

<sup>7</sup> <https://rd-alliance.github.io/metadata-directory/>

3. After the interview and the data description session, a document with RDM recommendations was prepared and sent to the researchers. In order to do so, we did a SWOT analysis based on the data collected in the interviews and metadata session. Such analysis was complemented with a survey of various RDM services and guidelines, which allowed a selection of recommendations according to the needs of the project. This document is not only a first point of assessment of the recommendations for the development of the DMP, but also a mechanism for benchmarking RDM services. As researchers are mostly developing their awareness, in some cases we choose to recommend more than one alternative for the same function. A good example is the proposal of a disciplinary repository, but also the suggestion to consult the re3data.org (Registry of Data Repositories), so that researchers can develop a broader understanding of available services. A selection of the Research Data Alliance recommendations has also been made to align our proposals with RDM good practices.
4. At a later point in time, we seek feedback regarding the proposed recommendations. A follow-up questionnaire was designed for researchers to assess whether the proposals were useful and helped to enrich their RDM knowledge. The purpose of this questionnaire was to evaluate each recommendation individually, but also to make a general assessment on their interest and of the perceived difficulty of implementing the recommendations.

As shown in Figure 2, these activities lead to the development of the DMP for the FrailSurvey project. Our workflow included different moments of evaluation by the researchers, one carried out through the assessment of the initial set of recommendations, while we also sought their feedback of a DMP version at the end of the process.

As shown in Figure 2, these activities lead to the development of the DMP for the FrailSurvey project. Our workflow included different moments of evaluation by the researchers.



**Figure 2.** Workflow for the DMP development

In the next sections we present the results from the interview and data description session carried out with the researchers that precedes the elaboration of the document with RDM recommendations.

## Steps in the Development of the DMP

Although the DMP is the gateway to formal data management for many researchers, its development is a somewhat complex task for those who are developing their skills in this area.

In this sense, available DMP models play a fundamental role, by presenting the topics that must be defined at each stage, according to funding agencies requirements. The DMP for the FrailSurvey project was instantiated in the DMPonline tool.<sup>8</sup>

### Assessment of RDM requirements in the FrailSurvey project

The FrailSurvey researchers interviewed recognized a lack of RDM knowledge and a lack of support for the different stages of the data lifecycle, which further motivated them to gain know-how to introduce data management practices in the project.

Data collection for the project is done through the FrailSurvey app, which consists of a set of questions relating to various dimensions to assess fragility. Questionnaire data pass directly from the app to the database, being organized in a spreadsheet. The researchers added that there was no form of identification of people and no restriction of participation. Hence, all data are accepted and stored in the database and, only later, for study and validation purposes, data from people under sixty-five are deleted.

As for access to data, only two people have access to the database, a researcher and the project manager. However, there is interest in sharing data in a data repository at the end of the project, with the intention of making the data accessible to every interested party, having in mind that areas such as health and the social sciences could benefit most from sharing this data.

Regarding the description and documentation accompanying the data, there is no creation of formal metadata at the time of data collection. In alternative, the researchers maintain a spreadsheet with the questions and answers obtained through the app. At the time of the interview there was no knowledge about data description models or available domain-specific standards. When asked about the necessary information to contextualize data, it was suggested a reference be made to the scientific areas to which the research may have interest.

Data management is something that researchers had not yet considered in previous projects. In this sense, researchers recognized that a recurring challenge is to maintain consistency in data documentation. With the execution of a DMP researchers have the expectation to facilitate data sharing and that this tool can contribute to improve data management during the project.

Based on the interview we performed a SWOT analysis as summarized in Table 1.

**Table 1. RDM SWOT analysis.**

Weaknesses and Risks	<ul style="list-style-type: none"> <li>• Manual deletion of the data, for study purposes, where the age is less than sixty-five;</li> <li>• Lack of description or documentation accompanying the data;</li> <li>• Researchers unaware of tools or models of data description;</li> <li>• The data set file presents little information (only the answer and question of the app);</li> </ul>
----------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<sup>8</sup> <https://dmponline.dcc.ac.uk/>

Weaknesses and Risks (continued)	<ul style="list-style-type: none"> <li>• The dataset stores all the data obtained through the app, although the scale for assessing frailty has been developed for people over sixty-five years old;</li> <li>• Inconsistency in projects that work with different databases. It does not present a description of the data that constitute it;</li> <li>• Lack of knowledge of data management practices may hinder applications for funding future projects.</li> </ul>
Strengths and Opportunities	<ul style="list-style-type: none"> <li>• New format – app as a new method of frailty evaluations;</li> <li>• Data collected through the app directly stored in the database;</li> <li>• Anonymization of participants;</li> <li>• The dataset is easy to understand;</li> <li>• Researchers willingness to share data and learn more about RDM;</li> <li>• Dataset will have no restrictions at time of sharing;</li> <li>• Frailty is a growing area of research;</li> <li>• Dataset has reuse potential in new research projects;</li> </ul>

## Assessment of RDM requirements in the FrailSurvey project

The data description session started by introducing FrailSurvey researchers to Dendro, with a brief demonstration of its features. The researchers were then asked to create a folder and upload their datasets. After this step it were explained in detail the choices that could be made in the vocabularies panel, together with an overview of the available descriptors in Dendro, with emphasis on the most appropriate for the domain and the type of data. Likewise, researchers were also introduced to Dublin Core concepts,<sup>9</sup> in order to enrich the metadata.

During the session the selection of descriptors was mostly up to the researchers. The exception was when feedback was required to explain the meaning of some metadata element, or when the researchers mentioned they were looking for a specific one.

A dataset about self-assessment of frailty, captured to make the validation of the survey based on the FrailSurvey application, was described in 20 minutes in a collaborative effort between the two researchers. The final metadata record includes 16 key-value pairs. Although the researchers' domain is the Life and Health Sciences, since the dataset was created via questionnaire it was recommended to start with DDI. The researchers briefly discussed the selection or rejection of some available descriptors. They quickly understood the concepts

<sup>9</sup> <http://dublincore.org>

represented in the DDI vocabulary, which made it clear that this was a suitable vocabulary for the creation of metadata and that the researchers were well acquainted with the concepts. The same is true for their assessment of Dublin Core elements, yet they were quicker to reject most concepts. The meaning of *Coverage* caused some doubt and was included in the metadata after a short explanation.

Halfway through the session the researchers asked if the metadata should be made in English or Portuguese which made it opportune to provide insight on the advantages of describing data in English for future reference. Another aspect that shows their awareness and commitment to the task was their intention to include all of the survey questions under the descriptor *Question*, however since the questionnaire was in a file, the file was uploaded upon recommendation, as a complementary document to the described dataset.

The final metadata record included descriptive information on the survey objects and benchmark on which it was based, while administrative metadata was captured to represent the people involved in the project and its target audience. Moreover, 6 descriptors were filled in in order to provide context metadata, such as the *Methodology* and the *Data Collection Methodology*. On top of that the researchers created metadata for the *Sample Size*, *Universe*, *Sampling Procedure*, *Instrument* and *Collection Mode*. The metadata also has one descriptor for semantic metadata (*Subject*), for technical metadata (*Format*) and Geospatial (*Coverage*) metadata.

No temporal metadata was recorded, although this information could be useful considering that the assessment of frailty may be linked to a specific economic and social context in time. Overall, the metadata created can support search and access to the data, has a balanced description with the use of standards that promote interoperability and sufficient study design information that may ease the reuse of data.

Data description was perceived as very useful by the researchers since it helps to systematize everything in a simple and more correct way. In their opinion the metadata was considered sufficient with no need for more information. As for the data description activity it was found slightly easy, fast and practical, yet a little discouraging.

## Selection and proposal of RDM resources

In the process of designing the DMP for the FrailSurvey project, a set of data management resources were surveyed. Mapping such resources is essential to select an adequate solution for each component of the DMP. Therefore, we have focused on resources tailored for the Health and Social Sciences and include them in the recommendations to researchers.

Table 2 summarizes the recommendations made to the researchers based on the insight gathered during the interview and data description session. It includes services and guidelines that were useful to define each proposal, when it applies, and also the specific resources for each. It should be noted that the proposed metadata models comprise vocabularies previously available in the Dendro platform. Thus, the data description session served to confirm the usefulness of the DDI and DC vocabularies for the FrailSurvey. These activities are not completely sequential, but also the result of incremental work.

The document was sent to researchers via email, and they were suggested to consult the different resources on their own to improve their RDM awareness, through autonomous analysis. The recommendations document was structured to be as self-explanatory as possible. Each recommendation was composed of a description and explanation of its importance, the definition of the resources to be implemented and, when applicable, a practical example based on the FrailSurvey project was also presented. The presentation of examples and additional information was fundamental to help researchers to understand and interpret each recommendation.

**Table 2. Recommendations proposed to researchers in preliminary document.**



Proposal	Services and guidelines	Resource
Metadata model adoption	Metadata Standards Catalog; RDA Metadata Standards WG	DDI + Dublin Core
Selection of data repository	Re3data.org	Zenodo; ICSPR
Complementary document to contextualize the dataset	-	-
Back-up plan	Health Research Board; Digital Curation Centre; OpenAIRE	-
Definition of data cleaning process	UK Data Service	UK Data Archive Research Data Lifecycle

The recommendations are related to several steps and elements of RDM. The first two recommendations portray two of the most important tasks to make data FAIR, the description supported by standards and data publication. As such it was recommended the adoption of a set of DDI elements to capture scientific-oriented metadata, complemented with Dublin Core metadata for descriptive and administrative purposes. The adoption of these standards ensures the desire for interoperability and help to promote the reusability of the data. For data publication it was recommended a domain-agnostic repository, to streamline publication, Zenodo<sup>10</sup> and a disciplinary, peer-reviewed alternative, the ICSPR repository.<sup>11</sup>

Another recommendation was the creation and development of a document that would accompany the dataset at the time of publication, providing a contextualization of the data of the project, as well as how the dataset is organized in order to facilitate interpretation and reuse by future stakeholders. The fourth recommendation suggested the definition and implementation of a backup plan, with preference being given to performing this task through automatic methods, considered by the Digital Curation Centre as the best strategy for performing backups. According to the Health Research Board Ireland,<sup>12</sup> data should be stored in two separate locations and that backups performed regularly to mitigate the risk of data loss. The document also advises that the dataset should not include redundant information.

In the researchers' opinion the metadata elements presented and their description were easy to understand and that this recommendation improved their awareness about available metadata standards, since these were previously unknown. In this case, they consider that it is important to have further contact with these resources. The same is true for the data repository suggestions, although one of the researchers does not know if exploring the repositories on his own is sufficient.

The researchers were fully informed about the need to elaborate on a complementary document to contextualize the data, the example provided being very useful to illustrate this recommendation since they were unaware of the usefulness of this type of practice to encourage data reuse. Likewise, the researchers considered the definition of a backup plan and a process to clean the data as useful and easy to understand. These two recommendations were found to be the most interesting to implement.

The adoption of a metadata model, the definition of a complementary document to contextualize data and of a backup plan were considered easy or very easy to implement, while

<sup>10</sup> <https://zenodo.org/>

<sup>11</sup> <https://www.icpsr.umich.edu/web/pages/>

<sup>12</sup> [https://dmponline.dcc.ac.uk/template\\_export/1638514350.pdf](https://dmponline.dcc.ac.uk/template_export/1638514350.pdf)

one of them considered that the selection of a suitable data repository and the data cleaning process were of moderate difficulty.

Overall, the FrailSurvey researchers considered that the recommendations document was easy to interpret and very useful to increase their awareness of different RDM practices, pointing that this easy interpretation was largely due to the practical examples associated with their project. In the next section we provide in more detail the DMP for the FrailSurvey, as an instance of the development approach outlined in this work.

## Developing a DMP for the FrailSurvey Project

Although the DMP is the gateway to formal data management for many researchers, its development is a somewhat complex task for those who are developing their skills in this area. Likewise, it can be a challenge even for data curators who may experience difficulties in defining what to implement and specify at each stage, when domain and project knowledge is not sound.

In this sense, the models for DMP development play a fundamental role, by presenting the topics that must be defined at each stage, according to funding agencies requirements. DMP building support platforms, such as DMPonline and DMPTool cover and integrate a large number of models, which makes them a good starting point for this activity. The model chosen for the FrailSurvey project DMP was the Digital Curation Centre model. The choice was mainly due to the fact that this model features general guidelines for research communities, which fits the needs of the project.

This model consists of a checklist that presents the main issues or themes that researchers should address in the preparation of the DMP. In each section the model presents a set of questions and examples to guide in the writing of the plan, and it is structured as follows: Data Collection; Documentation and Metadata; Ethics and Legal Compliance; Storage and Backup; Selection and Preservation; Data Sharing; Responsibilities and Resources.

The DMP for the FrailSurvey project was instantiated in the DMPonline tool, as depicted in Figure 3, with the corresponding seven sections.

1. Project data are observational and come from the responses of the app users to a set of questions from various areas such as health, social, psychological, socio-demographic, economic, physical, polymedication, and leisure. It was defined that the data collection is performed directly through the app.
2. At the time of collection data are not accompanied by any document describing them, so the creation of a "ReadMe" file is necessary when the data is shared and published. This document was thought of and defined with the objective of facilitating the reuse of the data of this project, since it will help other researchers to understand and interpret the data. It was also specified that, after being collected, the data would be described, possibly through tools such as Dendro, using the DDI and DC metadata standards, based on the results of the data description session.
3. The FrailSurvey project does not present any ethical problem regarding the data, since no personal data is collected, thus there is no possibility to identify the participants. The Copyright and Intellectual Property issue was addressed to define the institution that holds the rights over data.
4. Project data is automatically stored on the server in a robust database developed by institutional computer services. Access to the data is only possible through a dashboard, protected by credentials, which only a researcher and the person responsible for the project had. In this dashboard it is possible to download the file related to the dataset. It was defined the backups' periodicity, preferably through automatic methods, to be backed-up weekly or twice a week.

5. The data will be processed to eliminate the non-significant data, and the data that would remain would constitute the final dataset. After this task, the dataset is of great value in the short, medium and long term, since the research in the area of fragility is growing.
6. The dataset would be shared possibly through two research data repositories, Zenodo and through an institutional data repository, under the Creative Commons BY 4.0 license.
7. It was established that the responsibility for implementing the DMP would rest with the Principal Investigator, supported by a data trustee, and no additional resources would be required for its implementation.

Moreover, a set of specific guidelines were identified to comply with the FAIR principles, among others:

1. The assignment of a unique and persistent identified to the data;
2. The metadata needs to be accessible even if the data were no longer available;
3. Metadata would use formal, accessible, sharable and applicable language for knowledge representation;
4. The data must be shared with the associated license in a clear manner and must be associated with its origin.

## **Conclusion**

Despite the ever-increasing production of research data, and the introduction of new policies to promote access to them, RDM is still poorly widespread practice among research groups, particularly in small projects. Hence, the adoption of a DMP in research projects is not yet generalized. The paradigm is favourably changing but more localized cases will be needed to strengthen communities' confidence in investing their time and resources to RDM.

In this paper we described work conducive to the development of a DMP for a small research project. Two main objectives were achieved. One is the proposal of a collaborative approach for the development of a DMP, instantiated in the FrailSurvey project. The other, is a consequence of the engagement of researchers in RDM activities. Raising researchers' awareness is essential, not only for their personal growth, but also for them to increasingly feel prepared to collaborate with others in the development of DMPs in upcoming projects. The replication of this work with more research groups will likely promote the sharing of more FAIR data. Like data, DMPs should also be reusable as much as possible.

It should be considered that the DMP is often the first contact of researchers with RDM and therefore the proposals should not be very specialized, at risk of discouraging further engagement. The process of knowledge acquisition should be as agile as possible; therefore, our approach followed a principle of simplicity, fostering the contact of researchers with the services and tools that can guide the establishment of FAIR-compliant practices. It is an approach that also brings benefits to data curators and other information professionals, by playing the role of linking recommendations, resources, concepts and data management practices with research projects in different domains. The ability to relate resources to the needs of researchers is something we consider essential in institutions with limitations to implement services fully dedicated to RDM.

Although designed to be executed in person, many of the activities carried out in this work were done remotely. For instance, the recommendations document had to be sent by email to the researchers, who autonomously performed the analysis of the resources proposed. Nevertheless, the researchers have shown openness in acquiring RDM skills, namely how to prepare data for publication, at a more advanced stage of the project. For the FrailSurvey researchers the DMP will help to better structure other projects, in what they believe adds value to funding application and execution of other projects. In their opinion some actions are needed to increase awareness. In addition to lectures, seminars and workshops, face-to-face meetings end up being very useful to improve understanding and clarify their doubts. Researchers are interested in practical disciplinary examples and case studies, which reinforces the importance of also adopting this type of approach for pedagogical purposes.

## Acknowledgments

The authors thank European Union's Horizon 2020 project, Fair4Health, under grant agreement No 824666. The work was also supported by UID/MULTI/04378/2019 with funding from FCT/MCTES through national funds, and under the grant attributed to Luís Midão (SFRH/BD/137090/2018)

## References

- Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation* 8(2). doi:10.2218/ijdc.v8i2.263
- Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., ... Wouters, P. (2004). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal* 3. doi:10.2481/dsj.3.135
- Bishoff, C., & Johnston, L. (2015). Approaches to data sharing: An analysis of NSF data management plans from a large research university. *Journal of Librarianship and Scholarly Communication* 3(2). doi:10.7710/2162-3309.1231
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology* 63(6). doi:10.1002/asi.22634
- Carlson, J. (2010). The data curation profiles toolkit: Interview worksheet. *Data Curation Profiles Toolkit*. Paper 3. doi.10.5703/1288284315652
- Curry, R. G. (2016). Factors influencing research data reuse in the social sciences: An exploratory study. *International Journal of Digital Curation* 11(1). doi:10.2218/ijdc.v11i1.401
- European Commission (2016). *H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020*. Retrieved from [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- European Commission, Directorate-General for Research and Innovation, (2018). *Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data*, Publications Office. doi:10.2777/1524

- Green, P., Cairns, A., & White, H. (2019). Evaluating data management plans: Are they good and are they effective? In *Proceedings of the 40th IATUL Conference, Jun 23-27 2019*, Paper 3. Retrieved from <http://hdl.handle.net/20.500.11937/77695>
- Hey, A. J., Tansley, S., & Tolle, K. (2009) *The fourth paradigm: Data-intensive scientific discovery*. Redmond, WA: Microsoft Research.
- Hodson, S., Jones, S., Collins, S., Genova, F., Harrower, N., Mietchen, D., Petrauskaitė, R., & Wittenburg, P. (2018). *FAIR data action plan: Interim recommendations and actions from the European Commission Expert Group on FAIR data* (pp. 1-21). doi:10.5281/zenodo.1285290
- Jones, S. (2011). How to develop a data management and sharing plan. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Retrieved from <https://www.dcc.ac.uk/guidance/how-guides/develop-data-plan>
- Kennan, M. A., & Markauskaite, L. (2015). Research data management practices: A snapshot in time. *International Journal of Digital Curation* 10(2). doi:10.2218/ijdc.v10i2.329
- Kim, Y., & Stanton, J. M. (2016). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology* 67(4). doi:10.1002/asi.23424
- Ribeiro, C., Silva, J. R. da, Castro, J. A., Amorim, R. C., & Lopes, J. C. (2016). Projeto TAIL-Gestão de dados de investigação da produção ao depósito e à partilha (resultados preliminares). *Cadernos BAD*, 2. doi:10.48798/cadernosbad.1603
- Rocha da Silva, J. (2016). *Usage-driven application profile generation using ontologies*. PhD thesis. Faculdade de Engenharia da Universidade do Porto.
- Simms, S. R., & Jones, S. (2017). Next-generation data management plans: Global machine-actionable FAIR. *International Journal of Digital Curation* 12. doi:10.2218/ijdc.v12i1.513
- Smale, N., Unsworth, K., Denyer, G., Barr, D. (2018). The history, advocacy and efficacy of data management plans. *bioRxiv* 443499. doi:10.1101/443499
- Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., Parsons, M., Robinson, E., & Wyborn, L. (2019). Make all scientific data FAIR. *Nature* 570. doi:10.1038/d41586-019-01720-7
- Tenopir C., Allard S., Douglass K., Aydinoglu A. U., Wu L., Read E., Manoff, M., Frame, M. (2011). Data Sharing by scientists: Practices and perceptions, *PLoS ONE*, 6(6). doi:10.1371/journal.pone.0021101
- Tenopir C., Dalton E. D., Allard S., Frame M., Pjesivac I., Birch B., Pollock, D., Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE* 10(8). doi:10.1371/journal.pone.0134826
- Vines, T. H., Albert, A., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., Gilbert, K. J., Moore, J. S., Renaut, S., & Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology* 24(1). doi:10.1016/j.cub.2013.11.014

- Wicherts, J. M., Bakker, M. & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE* 6(11). e26828. doi:10.1371/journal.pone.0026828
- Wiley, C., & Kerby, E. (2018). Managing research data: Graduate student and postdoctoral researcher perspectives. *Issues in Science and Technology Librarianship* (Spring 2018) doi:10.5062/F4FN14FJ
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3(160018). doi:10.1038/sdata.2016.18
- Yoon, A. (2016). Red flags in data: Learning from failed data reuse experiences. *Proceedings of the Association for Information Science and Technology* 53(1). doi:10.1002/pr2.2016.14505301126
- Zenk-Möltgen, W., Akdeniz, E., Katsanidou, A., Naßhoven, V. & Balaban, E. (2018). Factors influencing the data sharing behavior of researchers in sociology and political science. *Journal of Documentation* 74(5). doi:10.1108/JD-09-2017-0126