FIOCRUZ

Data Management Plan: Elaboration Guide

AUTHORS:

Viviane Veiga
Simone Dib
Patricia Henning
Francisco José T. Nascimento
Paulo Guanaes
Erick Penedo
Cristina Guimarães

Data Management Plan: Elaboration Guide

Cataloguing at source

Oswaldo Cruz Foundation

Institute of Communication and Scientific and Technological Information in Health

D232 Data management plan [recurso eletrônico]: elaboration guide / Viviane Veiga, Simone Dib, Patricia Henning, Francisco José T. Nascimento, Paulo Guanaes, Erick Penedo, Cristina Guimarães. – Rio de Janeiro: FIOCRUZ, ICICT, Rede de Bibliotecas, 2022.
32 p. : il. color.

Bibliografia: p. 29-31.
Modo de acesso: World Wide Web.
ISBN: 978-65-87975-04-7.
Tradução de: Plano de gestão de dados de pesquisa: guia de elaboração.

1. Gestão de Dados de Pesquisa. 2. Plano de Gestão de Dados. 3. Dados de Pesquisa. 4. Ciência Aberta. I. Veiga, Viviane. II. Dib, Simone. III. Henning, Patricia. IV. Nascimento, Francisco José T. V. Guanaes, Paulo. VI. Penedo, Erick. VII. Guimarães, Cristina.

CDD 020

Ficha catalográfica elaborada por Regina Maria de Souza – CRB-7/7438, Biblioteca de Manguinhos / ICICT / FIOCRUZ – RJ.

Rio de Janeiro, , October 2022

First edition

TEAM

Oswaldo Cruz Foundation

President
Nísia Trindade Lima

Vice-Presidency of Environment, Health Care and Promotion
Patrícia Canto Ribeiro (Interim)

Vice-Presidency of Education, Information and Communication
Cristiani Vieira Machado

Vice-Presidency of Management and Institutional Development
Mario Santos Moreira

Vice-Presidency of Research and Biological Collections
Rodrigo Correa de Oliveira

Vice-Presidency of Health Production and Innovation
Marco Aurelio Krieger

The Office of the Chief of Staff
Juliano de Carvalho Lima

Institute of Communication and Scientific and Technological Information in Health

Director
Rodrigo Murtinho

Vice-Director of Information and Communication
Tania Cristina Pereira dos Santos

Vice-Director of Research
Mônica de Avelar Figueiredo Mafra Magalhães

Vice-Director of Education
Mel Bonfim

Vice-Director of Institutional Development
Leonardo El-Warrak

Chief of Staff
Claudenice Carvalho Girão

Research Data Management Plan - Dmp | Preparation Guide

Coordination
Institute of Scientific and Technological Communication in Health

Fiocruz Library Network
Laboratory of Scientific and Technological Information in Health

Authors

Viviane Veiga
Institute of Communication and Scientific and Technological Information in Health (Icict/Fiocruz)

Simone Dib
Institute of Communication and Scientific and Technological Information in Health (Icict/Fiocruz)

Patricia Henning
Federal University of the State of Rio de Janeiro (UNIRIO)

Francisco José T. Nascimento
Vice-Presidency of Education, Information and Communication (VPEIC/Fiocruz)

Paulo Guanaes
Joaquim Venâncio Polytechnic School of Health (EPSJV/Fiocruz)

Erick Penedo
Institute of Communication and Scientific and Technological Information in Health (Icict/Fiocruz)

Cristina Guimarães
Institute of Communication and Scientific and Technological Information in Health (Icict/Fiocruz)

Graphic project

Valéria de Sá - Icict/Fiocruz

Contributions

Research Data Management Plan Working Group | Fiocruz Open Science Forum

Funding source

Inova ICICT-Fiocruz/Oswaldo Cruz Foundation

TABLE OF CONTENTS

PREFACE

Every Brazilian researcher who needs a Data Management Plan should read this guide. It is independent of specific funder requirements. This guide was created within the partnerships between Fiocruz, GO FAIR Brazil Health, and GO TRAIN, and is a result of analysis and customization of international DMP resources to Brazilian needs. Specifically, it is based on the Science Europe guidelines, and templates used by popular tools for creating data management plans, such as DMPTool and DMPOnline. Furthermore, it pays special attention to the correct interpretation of legislation and ethical issues within the Brazilian context. Creating a DMP allows researchers to better plan their research because they need to make it explicit what data will be produced or reused, how and where it will be stored, and under what conditions and where it will be shared with others. This in turn fosters reproducibility and improves the visibility and uptake of research results. This is because others can have higher confidence in the results achieved when the data is well-managed and documented. Well written DMPs can also help in reducing the overall effort of managing data, as well as make it easier to fulfil funders' requirements. I hope this guide can help Brazilian researchers improve practices in research data management and will help them achieve high reproducibility of their research.

Tomasz Miksa, TU Wien, Group Coordinator RDA DMP Common Standard

This guide is an important contribution from ICICT/Fiocruz to the Brazilian health science community. It will serve as a useful and practical resource to help in the development of Data Management Plans, as well as an instrument for the much-desired internationalization of Brazilian research. I hope it will be used as a reference both nationally and internationally.

João Moreira, professor at UTwente

The publication of this guide represents another major step towards the implementation of Open Science policies. By contributing to better organization of data, it allows for concrete advances in research processes, collaborating with a cultural change in the modus operandi of science production and opening up the possibility of sharing data with other researchers. The launch of this guide, together with the Arca Dados repository and the FioDMP, follows in the footsteps of Fiocruz's successful Open Access policy and the Arca Institutional repository, which paved the way for the expansion of Open Science at Fiocruz.

Rodrigo Murtinho, Director of ICICT/Fiocruz

Fiocruz creating and publishing this guide is another contribution to the implementation of Open Science in Brazil. It is an instrument that will help all Brazilian researchers and research institutions to decipher and act effectively with the research data that is so important for scientific development itself. Organizing, providing access to, and preserving research data will allow Brazil to be among the countries that do science in a contemporary way i.e., in an open way.

Bianca Amaro, coordinator of the Brazilian Program of Open Science/IBICT

Fiocruz has once again taken the lead in promoting Open Science when developing this guide that will help in the preparation of documents for the planning of scientific research. For funding agencies, it will support the standardization of data management plans. For libraries, it will assist the researcher in the elaboration of their data management plans. Furthermore, for researchers, it will serve as a tool for planning and directing their research methodology.

Luana Salles, coordinator of GOFAIR Brazil, IBICT

Data planning and management are key for the development of all stages of research. In this sense, the guide will be an important reference to support Brazilian researchers in the construction of Data Management Plans in the formats currently required. It adds to the various actions to promote science that have been developed over the years as part of Fiocruz's commitment to the democratization of production and access to knowledge in health.

Cristiani Machado, Vice President of Education, Information and Communication at Fiocruz

INTRODUCTION

This Guide was created through a partnership between Fiocruz, GO FAIR Brasil Saúde and GO TRAIN under the coordination of the ICICT/ Research Data Management Plan Working Group of the Fiocruz Open Science Forum.

At present, several research funders require a Data Management Plan (DMP) aligned with the FAIR principles (Findable, Accessible, Interoperable and Reusable) as part of project submission. In turn, Teaching and Research Institutions are implementing campaigns to raise awareness among the academic community about the importance of DMPs and encourage their use. However, this activity places new demands on researchers who very often feel insecure at the time of preparing their plan.

Without wanting to exhaust the topic in question but based on the main doubts arising in the preparation of DMPs in Brazil, this guide aims to aid researchers, teachers, and students of postgraduate courses in this practice. It was conceived to summarize international initiatives for data management, for use in any area of study, with special attention paid to the field of health. All sections were developed based on the analysis, interpretation, and translation of the guidelines on filling in DMPs contained in the templates of the main tools used to prepare DMPs in the United States and Europe, such as DMPTol and DMPOnline, and the guidelines set out in the Practical Guide to the International Alignment of Research Data Management - Extended Edition 2021.

It is observed that, in practice, all the templates of the DMPs analyzed tend to follow the Science Europe format, with minor variations that have been considered in this guide.

INTRODUCTION

In the section on legislation and code of conduct, the normative and ethical framework necessary in Brazilian research practice was considered.

Not all these sections are necessarily always required in a DMP and there may be a slight variation between institutions. However, in view of these small variations and the difficulties researchers have in responding to them, we have developed an overview of the requirements that may be made in a DMP, thus opening a range of possibilities.

We draw the researchers' attention to some important points to remember.

This Guide is the first step to stimulate, in a practical way, the culture of data management within research and education institutions. We hope that the information contained herein can help you in developing your Data Management Plan

Make effective use of it!

1

DATA DESCRIPTION, DATA COLLECTION AND/OR REUSE

1.1 – WHAT DATA WILL BE COLLECTED OR CREATED

It is important that all datasets are described in a DMP prior to research. This question is related to the description of the origin, type, volume, and format of the data. To answer it, describe:

Regarding the origin of the data: describe whether the research data are observational (obtained through direct observation), experimental (coming from experiments, usually conducted on laboratory benches) or computational (data generated from computer models or simulations). In the case of reuse of data that have been collected for other purposes describe whether it is governmental, administrative, social media, etc.

Regarding the type of data: inform the type(s) of data that will be generated/collected: e.g., numerical (spreadsheets), textual (documents), image, audio, video, etc. If you are working with more than one type, specify all of them.

Regarding the volume of data: please enter the approximate volume of data that will be generated/collected in Megabyte (MB), Gigabyte (GB) or Terabyte (TB). Specify the ramifications related to the volume of data in terms of storage space, location/access, and preservation. Furthermore, whether any cost will be incurred for obtaining and/or storing the data.

Regarding the format of the data: please provide details of the formats of the files containing the data. For example, whether they are: txt, csv, tif, pdf, xls, doc, rdf or another

IMPORTANT TO CONSIDER:

Some formats are ideal for preserving data over the long term.For example, non-proprietary formats, such as text ('.txt') or ('.csv') are considered easy to preserve.

The UK Data Service provides a useful table of file formats for various types of data.

DATA DESCRIPTION, DATA COLLECTION AND/OR REUSE

type. Indicate which community data format standards (if any) will be used; industry standard formats, such as dicom, NIFTI; European data format such as edf; or BrainVision data format such as eeg, vhdr, vmrk.

Specify if there is data to be converted into other formats.

Justify the choice of formats. It may have been based on the experience of the research team, on the preference for open formats, on standards accepted by the data center, or on widespread use by a particular community.

Proprietary file formats, which require specialized software or hardware to use, are not recommended for preservation purposes, but may be necessary for certain data collection or analysis methods. Use open file formats whenever possible.

REMEMBER:

Give preference to the use of standardized, interchangeable, or open formats, which ensure the usability of the data in the long term.

1.2 - HOW WILL THE DATA BE COLLECTED OR PRODCUED AND HOW CAN IT BE REUSED?

This question relates to the standards and/or methodologies that will be adopted for the collection or production of data. To answer it, please describe:

a) what methodology will be adopted for the collection/ production of the data;

b) how files and folders will be structured and named;

c) whether version control of the data will be performed. If positive, how;

d) what quality assurance procedures will be adopted.

DATA DESCRIPTION, DATA COLLECTION AND/OR REUSE

REMEMBER:

Consistent and well-organized research data will be easier to find, understand and reuse.

This will require:

(a) indicating whether the data are owned by third parties (data collected from institutional websites, databases, or reused data from other researchers) or if the researcher will generate them. Please describe their provenance in detail;

For example: Who generated or collected them? How were they processed? Have they been published before?

(b) assessing, if applicable, how the data generated could be integrated with third-party data and whether the possibility of reusing them truly exists. Describe the step-by-step approach taken, considering the content, coverage, and type of data. For example, spreadsheet data, experimental measurements, models, software, audio-visual data, physical samples etc;

(c) describing how the data will be organized during the project, mentioning, for example, naming conventions, version control and folder structures;

(d) describing how the consistency and quality of data collection will be controlled and documented. This may include calibration processes, repeating samples or measurements, standardized data capture or recording, validation of data entry with peer review, or representation with controlled vocabularies;

(e) explaining which software will be used for the data collected or produced.

2

METADATA, DOCUMENTATION AND DATA QUALITY

2.1 – WHAT METADATA STANDARD WILL ACCOMPANY THE DATASETS?

Metadata is data about data. It contains the description (of datasets of the same nature) necessary for the data to be read and interpreted in the future. It is of utmost importance not only for preserving memory, but also to help with the interoperability and reproducibility of data. To answer this question, describe:

REMEMBER:

Researchers are strongly encouraged to adopt metadata standards from the community they are working in. The Research Data Alliance (RDA) initiative has developed a directory of metadata research standards, currently maintained, and updated by the Digital Curation Center (DCC). It is called Disciplinary Metadata.

(a) depending on the nature of the datasets, which metadata standards could be used. For example: DDI, TEI, EML, MARC, CMDI, Dublin Core, Darwin Core;

(b) if your dataset is stored in a repository, please provide the standard used by the platform, considering that some repositories allow the use of more than one metadata standard. Whenever possible, give preference to the specific metadata standard in your area of knowledge.

For the metadata description to be aligned with the FAIR principles, the following requirements must be met:

(a) be rich and comprehensive enough to clearly describe the data set;

(b) contain a persistent identifier for both the dataset and for the metadata;

(c) remain stored, in the long term, even when the data is no longer available;

(d) describe the provenance of the data and follow the community standard (if any);

METADATA, DOCUMENTATION AND DATA QUALITY

(e) be stored, together with the dataset, in a trusted repository.d) descreverem a proveniência dos dados e seguir o padrão da comunidade (quando houver);

2.2 - WHAT TYPES OF DOCUMENTATION WILL ACCOMPANY THE DATASETS?

To help other researchers and users to find, understand and reuse your datasets, consider creating a “readme” text file or the like. This file could be used to report additional information, such as: the methodology used to collect the data; analytical information; definitions of variables; units of measurement; any assumptions made; the format and type of data file; and the software used to collect and/or process the data.

REMEMBER:

The FAIRSharing portal offers a database for searching standards, repositories, and policies. https://fairsharing.org/

3

LEGAL AND ETHICAL REQUIREMENTS AND CODES OF CONDUCT

3.1 – HOW WILL ETHICAL ISSUES AND CODES OF CONDUCT BE MANAGED?

While codes of conduct may vary according to the culture of an academic community, questions about the confidentiality of personal data and the ethical nature of research are based on more universal principles and values. To meet research criteria, the researcher needs to answer the following questions:

a) how will the privacy of research participants be protected where necessary? A common example is through anonymization;

b) how will confidential data be handled to ensure that it is stored and transferred securely?

c) how will the rights of third parties be respected, when data from other sources are used or reused, i.e., that were not produced exclusively in the research?

d) what will be the access levels (open, embargoed, restricted, or closed) assigned to the dataset?

e) when the research presents the need for submission to ethical analysis, as in the case of research involving human beings or the use of animals, the management of such issues should consider: the submission of the project to departmental or institutional ethics committees; the preparation of Free and Informed Consent Terms, Record of Informed Consent and Free Informed Assent Terms where appropriate. Furthermore, the anonymization or pseudonymization of personal data, especially sensitive data, among other procedures.

REMEMBER:

Ethical issues influence not only the collection and storage of data, but also accessing, sharing, and reusing data: -

- Who will be able to access and/or use them?

- Will there be a possibility to reuse or not?

- How long will they be kept?

It is therefore important for the researcher to know some basic legal and ethical regulations so that their research is feasible.

LEGAL AND ETHICAL
REQUIREMENTS
AND CODES OF CONDUCT

The elaboration of the Free and Informed Consent Terms, Record of Informed Consent and Free Informed Assent Terms must follow the guidelines and regulations in force, describing, when applicable, the possibility or not of sharing and reuse of data, making the access conditions clear.

ETHICS AND PRIVACY

If your research does not correspond to any legal or ethical questions, this information should be clearly described in your Data Management Plan.

Researchers conducting studies involving human subjects should be aware of the regulatory guidelines and standards for research of this nature, as well as issues related to data management when preparing the study participant's record of consent or assent.

This participant needs to be informed about procedures to protect their privacy and the confidentiality of the data.

Where possible, re-use of data should be consented to by the by the participant, considering anonymization or pseudonymization as far as is practicable.

Ethical issues may define how the data should be stored and transferred, who will be able to see/use it and for how long it will be kept. Demonstrate that you are aware of this and that you have planned accordingly. Always consult your institution's Research Ethics Committee linked to the National Research Ethics Committee.

Make sure that when dealing with the processing of personal data, the laws protecting this type of data are complied with. One such example is the General Law on the Protection of Personal Data. Observe these procedures:

a) obtain, where necessary, informed consent for the preservation and/or sharing of personal data and for all

data processing operations. As a rule, consent for the use of data is limited to the specific research being conducted;

b) consider the anonymization of personal data for preservation and/or sharing (anonymized data is not personal data, as provided for in the General Law on the Protection of Personal Data);

c) in general terms, pseudonymization is a treatment which involves the extraction of additional information from personal data; this information, when reintegrated to the pseudonymized data, allows the identification of the holder of the personal information. Therefore, this additional information must be stored separately in a secure and access-controlled environment;

d) describe whether there is a controlled access procedure for authorized users of personal data.

If you plan to deposit and share the data on digital platforms, it is important to state in the consent instrument that the anonymized data will be made available in institutional or disciplinary data repositories, even if the access regime is restricted or closed.

The record of consents should be deposited together with the research data in a secure and controlled environment, such as in a data repository.

In the case of research involving the use of animals, approval from the Ethics Committee on the Use of Animals is required.

Minutes from the research approval process or the consolidated opinion of the Research Ethics Committee should be included in the dataset to be preserved.

It is important to emphasize that it will be up to the Research Ethics Committee to analyze the possibilities of sharing, reuse and, when necessary, anonymization or pseudo anonymization.

REMEMBER:

Consider the pseudonymization of personal data, especially sensitive data (the main difference from anonymization is that pseudonymization is reversible. Therefore, pseudonymized data are considered personal data in the General Law on the Protection of Personal Data)

LEGAL AND ETHICAL
REQUIREMENTS
AND CODES OF CONDUCT

3.2 – HOW WILL THE LEGAL ISSUES OF INTELLECTUAL PROPERTY RIGHTS (IPRS) AND AUTHORSHIP BE MANAGED? WHAT LEGISLATION IS APPLICABLE?

This question is related to issues surrounding intellectual property, which protects copyright (economic and moral of the holder of these rights) and industrial property (patents, industrial secrets, among others). Issues concerning these two areas are the most likely to arise within the research cycle or during the process of opening, accessing, sharing, and reusing research data.

To answer this question, describe:

a) who will be the owner of the intellectual property rights of the of the data resulting from the research, i.e., who will hold the rights to control access to all existing data, as well as to new data that will be generated. For projects with one or more partners, the terms and conditions agreed based on IPRs should be included in the contract or consortium agreement;

b) what access conditions apply to the data; whether the data will be openly accessible or with access restrictions. In this case, describe what the restrictions are. State what the license(s) for use and reuse will be. Specify the legal rights and limitations of users;

c) any necessary restrictions on access and sharing of data. For example, to protect proprietary, confidential, strategic, patentable, or otherwise protectable data;

d) whether foreseen IPRs are in any way affected. An example would be the sui generis right within the European Database Protection Directive on database protection. If so, please specify which IPRs and how they will be addressed;

e) make sure to cover intellectual property issues in the

LEMBRE-SE:

Please refer to the ICPSR procedures on data confidentiality. Observe norms set by HIPAA on data protection and data privacy in health research.

LEGAL AND ETHICAL
REQUIREMENTS
AND CODES OF CONDUCT

contract or consortium agreement, to control access to data from projects with multiple partners and data owners. Consider the funder, the institutional and departmental policy, or the partner group;

f) also consider permissions to reuse third-party data and any necessary restrictions on data sharing.

If the research generates confidential data, it is necessary to describe how it will be overseen, the access regime and how it will be securely transferred.

REMEMBER:

in the case of partnerships with foreign institutions, the parties must agree on the legislation applicable to the agreement or contract according to their convenience, opportunity, and institutional interest, respecting the national laws of all partners. The parties must also choose the forum for resolving any disputes.

Thus, the terms and conditions of the agreement or contract must not contravene these frameworks. This calls for caution regarding the possible international transfer of data, which in Brazil, is unfeasible in some cases.

ACCESS TO INFORMATION LAW AND CONFIDENTIAL INFORMATION

Data from research conducted by public bodies, or those conducted with public resources, are, as a rule, accessible to the public. The control of access must be based on an equivalent to the Access to Information Law

The Access to Information Law sets out exceptions to the right of access to information, for example, as in the cases of industrial property, tax secrecy, business secrecy, professional secrecy, control of national biodiversity, conflict of interests, security of society and the State, national strategic interest, among others.

Whenever the research involves sensitive objects, rights and values, the researcher should be aware of the possibility of the occurrence of a legal hypothesis of confidentiality.

LEGAL AND ETHICAL
REQUIREMENTS
AND CODES OF CONDUCT

There are many legal hypotheses of secrecy existing in the Brazilian legal system. Here we indicate only a limited list. For this reason, it is important to have the advice of a trained professional to guide on the identification of these incidences and the adoption of the most appropriate data access regime.

For more information, please access the Oswaldo Cruz Foundation Procedure for classification and treatment of confidential information.

REMEMBER:

Legal and ethical requirements are complex and controversial, so do not be afraid to seek support from Research Ethics Committees and the legal advisors of your respective research organizations.

In the case of research funded with public funds, it is necessary to be aware of:

(1) the incidence of the Access to Information Law;

(2) the National Open Data Policies

As far as they provide for the access and sharing of data.

4

STORAGE AND BACKUP DURING AND AFTER THE RESEARCH PROCESS

4.1 – HOW WILL DATASETS AND METADATA BE BACKED UP AND STORED DURING AND AFTER THE RESEARCH PROCESS?

Planning how research data will be stored and backed up during and after the research is fundamental to ensure the security and integrity of the data. For this reason, it is important that all datasets are stored in an appropriate location, with a strict backup policy from the beginning of collection. To answer this question, it is necessary:

LEMBRE-SE:

Only storing data on laptops, computer hard drives or storage devices is very risky.

For this reason, robust storage managed by the Information Technology (IT) team of the institution is recommended.

a) to describe where the datasets will be temporarily stored before being selected and deposited for preservation purposes, e.g., institutional server, on the researcher's own computer or in cloud services;

b) indicate whether the researcher has sufficient space for temporary storage of their data or whether resources for additional services will be needed;

c) report on how data will be backed up: how often, in which locations, how many copies will be made and who will be responsible for this procedure. Storing data in at least two separate locations is recommended;

d) describe how data will be recovered in the event of an incident.

STORAGE AND BACKUP DURING AND AFTER THE RESEARCH PROCESS

REMEMBER:

If you choose to use a third-party service you must ensure that it does not conflict with any funder and/or institutional, departmental or group policies. This could be in terms of the legal jurisdiction in which the data is held or the protection of confidential data.

Note the main risks and whether there is any institutional data security policy in place.

Proper storage and backup not only help to protect research data from catastrophic loss (due to failures of hardware, software, viruses, hackers, natural disasters, human error, etc.),

but also enable appropriate access by current and future researchers.

4.2 – HOW WILL DATA ACCESS AND SECURITY BE MANAGED?

This question relates to access management and data security. To answer it, it is necessary to describe:

a) what risks exist and how they will be managed and recovered in the event of an incident;

b) who will have access to the data during the research and how control will be managed, especially in collaborative partnerships;

c) how the collaborators' ability to securely access data will be ensured;

d) if creating or collecting data in the field, how the secure transfer will be ensured;

e) how the protection/security of data will be assessed, particularly if it is confidential and contains personal and political information or industrial secrets;

f) whether data will need to be encrypted and how this will be done to ensure that it is not accessible to people outside the project. Check if the institution offers this type of service and how it can guide you;

g) how the data will be accessed at the end of the research. Choosing a trusted repository that assigns a persistent identifier (such as a DOI, Handle etc.) to your dataset is recommended. This will ensure stable access and will enable retrieval via various discovery tools.

REMEMBER:

If the data is confidential or sensitive (e.g., personal data not yet in the public domain, political data, sexual orientation information or trade secrets), you should outline any appropriate security measures and observe any formal standards to be met, such as ISO 27001 on information security management.

5

SELECTION AND PRESERVATION

5.1 – WHICH DATA WILL BE SELECTED FOR RETENTION, SHARING AND/OR LONG-TERM PRESERVATION?

It is important that datasets are selected for long-term preservation. To answer this question, describe:

(a) which datasets should be retained for preservation or discarded;

(b) the criteria adopted, considering the institution's preservation policy (if any), and the contractual, legal, or regulatory requirements.

REMEMBER:

Define which data to keep and for how long, based on obligations to retain certain data.

For example: the potential for reuse, what is economically viable to maintain, and the effort required to prepare the data for sharing and/or preservation, such as changing file formats.

5.2 – WHAT IS THE PLANNING FOR THE LONG-TERM PRESERVATION OF DATASETS?

This question is related to the preservation of datasets in the long term. To answer it, please describe:

a) where the data will be stored. Provide the name and address of the place;

b) how much work time is required to prepare the data for preservation;

c) how long the data will be retained for;

d) what resources will be needed to ensure that datasets remain usable in the future.

SELECTION AND PRESERVATION

REMEMBER:

The choice of data preservation will depend on the potential for reuse and the long-term importance of the datasets, as well as whether the researcher has obligations to funders or collaborators to retain the data or not.

When converted from one format to another the preservation files may lose information (for example, converting from an uncompressed TIFF file to a compressed JPG file).

For this reason, the changes in file formats must be documented.

The need to preserve data in the short term (for peer review purposes) or for the long term (for data of lasting value) will influence the choice of data repository. Search tools such as Re3data or Plos One are useful for finding an appropriate thematic repository for your data if there is no data repository in your institution.

Check if your institution provides a repository for research data and/or a policy for the sharing of research data.

We recommend that data, whenever possible, be deposited in the repository of the institution to which the researcher is linked or in another data repository under the responsibility of a Brazilian public institution. Links to data records should in some cases be replicated in thematic repositories within the research area, or in international funders repositories. It is important to ensure, whenever possible, that the data are physically deposited on Brazilian soil, under Brazilian law.

Depending on the type of research, it may be desirable to preserve all versions of the data (raw, processed, analyzed, final), but in some research it may be preferable to keep only selected or final data.

As not all repositories offer long-term preservation options, consulting the repository's policies before deciding where to deposit is recommended.

6

DATA SHARING AND REUSE

6.1 – HOW AND WHEN WILL THE DATA BE SHARED?

This question is related to sharing to enable data reuse. To answer it, please describe:

a) what data will be shared, e.g., raw, processed, analyzed and in what form;

b) with whom the datasets may be shared and under what conditions;

c) when the data will be made available for sharing;

d) how the data will be shared. The methods used will depend on a number of factors, such as type, size, complexity, and sensitivity of the data;

e) what type of license will be used for the datasets;

f) what are the restrictions for reuse of collected data, if any?

If potential users need specific tools to access and reuse the data, consider the sustainability of the software needed to access the data.

REMEMBER

Where there is potential for data reuse, you should adopt standards and formats that facilitate reuse and guarantee that the appropriate metadata is available online so that your data can be discovered. Persistent identifiers should be applied so that that the data can be found in a reliable and efficient way. They also help the tracking of citations and reuse.

DATA SHARING AND REUSE

6.2 – ARE THERE POSSIBLE RESTRICTIONS ON SHARING THE DATASET?

This question is related to existing restrictions for the sharing of data. It may be the case that, for example, the dataset needs to be embargoed for some reason. If there is no restriction, please advise. To answer this question, describe:

(a) whether there is any need to embargo the dataset. If so, please list the reasons, e.g., publication, protection of intellectual property - such as seeking patent protection, contractual obligations with potential partners, or other obligations undertaken etc;

(b) what action will be taken to overcome or minimize the embargo restrictions;

(c) who may use the data. If it is necessary to restrict access to certain communities or enforce a data sharing agreement, describe how and why;

(d) any expected difficulties regarding the sharing of data with recognized long-term value, together with the causes and potential measures to overcome them. For example, restrictions may be due to confidentiality, lack of consent agreements or IPRs.

REMEMBER:

Depositing the dataset in the repository does not mean that the data will be open. It is possible to leave data with closed or restricted access within the repository or offer it only on request.

Please refer to the policies and procedures of your institution and the repository in which the datasets will be deposited.

Consider possible strategies for minimizing sharing restrictions. These may include anonymizing, obtaining participant consent for the sharing of data and copyright permissions, and agreeing to a limited embargo period.

RESPONSIBILITIES AND FINANCIAL RESOURCES

7.1 – WHO WILL BE RESPONSIBLE FOR DATA MANAGEMENT?

This question is related to defining responsibilities for the management of datasets. To answer it, please describe:

a) who will be responsible for implementing the DMP and for ensuring that it is reviewed and updated. Please indicate the name and contact details of the person responsible;

b) who will be responsible for each data management activity. For example: data capture, metadata production, data quality, storage and backup archiving and data sharing;

c) how responsibilities will be divided between the partner initiatives in collaborative research projects;

d) whether data ownership and responsibilities for managing research data will be part of any consortium agreement or contract signed between the partners;

e) the responsibilities for coordinating data management between partners in the case of collaborative projects.

REMEMBER:

Consider regular updates to the DMP. At the beginning of the project the lead researcher should identify all the people who will have responsibilities for data management tasks during and after the project.

RESPONSIBILITIES AND FINANCIAL RESOURCES

7.2 – WHAT RESOURCES WILL BE REQUIRED TO DELIVER THE PLAN?

This question is related to the human, financial and technological resources that will be employed during data management and the delivery of the DMP. To answer it, describe:

a) whether any additional specialization or training for existing staff is required. Report any relevant technical expertise, support and training that are likely to be required and how to acquire them;

b) whether the data repository will charge for data storage and/or curation;

c) whether additional resources will be needed to prepare the data for the deposit or to meet any requirements from the data repositories;

d) what resources will be needed to implement your data management plan. Also describe what will be the overall estimated cost.

This estimate should include data management costs expected during the project as well as those required for long-term data support after project completion.

REMEMBER:

OpenAIRE has developed a tool to help researchers estimate the costs related to data management.

8

FINALIZED MY DATA MANAGEMENT PLAN, NOW WHAT?

FIRST OF ALL, CONGRATULATIONS ON PREPARING YOUR PLAN!!!

The next step is the deposit of the plan, preferably in a trusted data repository, which generates a persistent identifier such as a Digital Object Identifier (DOI). This identifier facilitates the DMPs' retrieval and citation over time and as such is one of the requirements for alignment with the FAIR principles.

Some systems for generating data management plans such as Fiocruz's system called FioDMP, have interoperability with a repository for depositing and, if the user wishes, automatic DOI generation. This deposit can be made in storage or sharing format. In the storage format, the record is allocated in a restricted area which can only be accessed and viewed by the researcher. In the sharing format, the DMP metadata record will be visible, but access to the DMP may be closed, embargoed or open, as decided by the researcher.

It is important to remember that the dataset deposit should be held in your institution's data repository where possible. If your institution does not have a repository and it is not possible to deposit in another national data repository, depositing in a free international repository such as Zenodo is recommended or the identification of a thematic repository, as already addressed in this Guide (section 5 - Selection and Preservation). At the time of sharing your

REMEMBER:

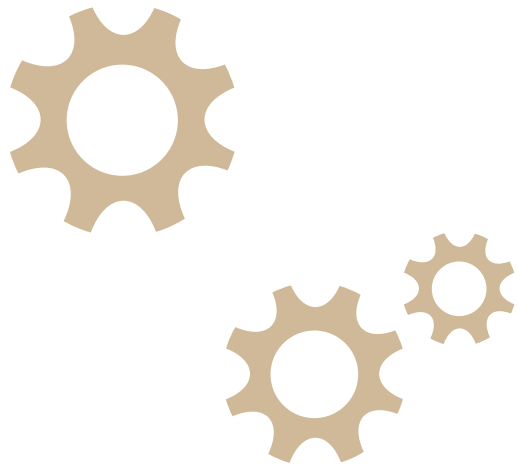
The FAIR principles (acronym for Findable, Accessible, Interoperable and Reusable) are the guiding elements of the research management process.

FINALIZED MY DATA MANAGEMENT PLAN, NOW WHAT?

dataset in a trusted data repository, you should include in the metadata description the “persistent identifier” of your DMP and/or deposit it with the dataset.

REMEMBER: The FAIR principles (acronym for Findable, Accessible, Interoperable and Reusable) are the guiding elements of the research management process.

GREAT DATA MANAGEMENT AND SUCCESS IN YOUR RESEARCH!



9

SOURCES CONSULTED

1 - DATA MANAGEMENT PLANS - DMPTool is a free and open-source online system that helps researchers create their data management plans (DMPs). This is a service of the University of California Digital Library. Available at: https://dmptool.org/ Accessed: 01 Sep. 2022.

2 - DATA MANAGEMENT PLANS - DMPOnline is a tool that helps the researcher create, review, and share their Data Management Plan, to meet institutional and research funding agencies' requirements. It is a service offered by the Digital Curation Center (DCC). It is used in Europe, predominantly in universities in the United Kingdom. Available at: https://dmponline.dcc.ac.uk/ Accessed on: 01 Sep. 2022.

3 - SCIENCE EUROPE - PRACTICAL GUIDE TO THE INTERNATIONAL ALIGNMENT OF RESEARCH DATA MANAGEMENT - EXTENDED EDITION, 2021. This is a Data Management Plan (DMP) guidance document for organizations, scientific communities, as well as individual researchers. Available at: https://bit.ly/3MVIbz9 Accessed on: 01 Sep. 2022.

4 - UK DATA SERVICE - RECOMMENDED FORMATS, formats recommended by the UK Data Service. It presents a table with the guidelines on file formats recommended and accepted by the institution for the sharing, reuse, and preservation of data. It may be necessary to convert data files to a proprietary format for preservation. Available at: https://ukdataservice.ac.uk/learning-hub/research-data-management/ Accessed on: 01 Sep. 2022.

5 - DISCIPLINARY METADATA - DIGITAL CURATION CENTER (DCC). This is a service offered by the DCC that connects disciplinary metadata, including tools to implement the standards. For those disciplines that have not yet established a subject-specific metadata standard, the General Research Data section contains links to information about broader metadata standards that have been adapted to meet the needs of research data. Available at: https://www.dcc.ac.uk/guidance/standards/metadata Accessed: 01 Sep. 2022.

6 - FAIRSHARING.ORG is a service of interest to producers, consumers of standards, databases, repositories, librarians, journal publishers, funding agencies and data management policy makers. In relation to standards, it offers various ontologies, metadata schemas and terminologies. It lists some data repositories and databases, as well as policies for data preservation, management and sharing from international funding, regulatory and journal agencies. Available at: https://fairsharing.org/ Accessed on: 01 Sep. 2022.

7 - APPROACH TO CONFIDENTIALITY - INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH (ICPSR). The ICPSR is concerned with risks in the dissemination of data. Once data is deposited with the ICPSR, the team uses rigorous procedures to protect the confidentiality of the individuals and organizations whose personal information may be part of the data collection archived. Available at: https://bit.ly/3EFyUJy Accessed on: 01 Sep. 2022.

SOURCES CONSULTED

8 - HEALTH INFORMATION PRIVACY. The HIPAA Privacy Rule establishes the conditions under which protected health information may be used or disclosed by research-oriented institutions. With respect to research, it protects the privacy of individually identifiable health information while ensuring that researchers continue to have access to medical information needed to conduct vital research. Available at: https://www.hhs.gov/hipaa/for-professionals/special-topics/research/index.html Accessed on: 01 Sep. 2022.

9 - ACCESS TO INFORMATION LAW - LAW NO. 12.527, FROM NOVEMBER 18, 2011. Art. 1 of Access to Information Law provides for the procedures to be observed by the Union, States, Federal District and Municipalities, to guarantee the information provided for in item XXXIII of art. 5, in item II of §3 of art. 37 and in §2 of art. 216 of the Federal Constitution. Available at: https://bit.ly/3gtTsKO Accessed on: 01 Sep. 2022.

10 - NATIONAL OPEN DATA POLICIES (PNDA) - BRAZILIAN OPEN DATA PORTAL. The open data policy of the Federal Executive Branch regulates and guides the publication of open government data by the bodies and entities of the Federal Government. This includes public agencies, authorities, and public foundations. Available at: https://bit.ly/3VLvmLY Accessed on: 01 Sep. 2022.

11 - PROCEDURE FOR CLASSIFICATION AND TREATMENT OF THE OSWALDO CRUZ FOUNDATIONS' CONFIDENTIAL INFORMATION. This Procedure for classification and treatment of confidential information at Fiocruz aims to comply with the provisions of art. 44 of Dec. 7.724/2012, which states that "the authorities of the Federal Executive Branch shall adopt the necessary measures so that the personnel subordinate to them know the rules and observe the security measures and procedures for the treatment of information classified as having a degree of secrecy". Available at: http://bit.ly/3zP2hFd Accessed on: 01 Sept. 2022.

12 - RE3DATA.ORG - REGISTRY OF RESEARCH DATA REPOSITORIES. Re3data offers a service that lists research data repositories from several countries, from different disciplinary areas. It fosters a culture of sharing, access, and better visibility of research data. The service went live in fall 2012 and was funded by the German Research Foundation (DFG). Available at: https://www.re3data.org/ Accessed on: 01 Sept. 2022.

13 - RECOMMENDED REPOSITORIES - PLOS ONE is a non-profit open access publisher that enables researchers to publish in open access, especially in the medical field. It provides a list of trusted repositories classified by area and type. Available at: https://bit.ly/3sehTyE Accessed on: 01 Sep. 2022.

14 - THE COST OF DATA MANAGEMENT - OPENAIRE. Information on how to calculate the cost of data management; how to use this costing tool and estimate the costs of managing and sharing data is offered by the OPENAIRE initiative: Available at: https://bit.ly/3EJX5GS Accessed on: 01 Sep. 2022.

15 - ZENODO is a data and information repository that is at the forefront of the open access and open data movements in Europe. It is a project led by the OpenAIRE initiative, launched in May 2013 to support open science. Available at: https://zenodo.org/ Accessed on: 01 Sep. 2022.

16 - DIGITAL OBJECT IDENTIFIER (DOI). DOI stands for Digital Object Identifier. It is comprised of a unique code that is presented in the form of a link, to identify and retrieve digital objects. It can be assigned to a publication, as well as to a set of data that are available on the internet. Available at: https://www.doi.org/ Accessed on: 01 Sep. 2022.

SOURCES CONSULTED

17 - FIODMP is the new platform for the preparation of a Research Data Management Plan created by ICICT/Fiocruz to serve the scientific community of Fiocruz and the broader scientific community of the health area. Available at: https:// fiopgd.icict.fiocruz.br/ Accessed on: September 1, 2022.

18 - FINDABLE, ACCESSIBLE, INTEROPERABLE AND REUSABLE - FAIR. The FAIR principles, an acronym for Findable, Accessible, Interoperable and Reusable, were consolidated in 2017, when the European Commission started to require the implementation of a data management plan based on these principles for projects that it funded. Since then, these principles have been guiding the discovery, access, interoperability, sharing and reuse of research data. Available at: https://www.nature.com/articles/sdata201618 Accessed on: 01 Sep. 2022.

19 - VEIGA, VIVIANE; HENNING, PATRÍCIA; DIB, SIMONE FAURY; PENEDO, ERICK; LIMA, JEFFERSON DA COSTA; SILVA, LUIS OLAVO BONINO DA; PIRES, LUÍS FERREIRA.
FAIR DATA MANAGEMENT PLAN: a proposal for Fiocruz. LIINC EM REVISTA , v. 15, p. 275-286, 2019. Available at: https://revista.ibict.br/liinc/article/view/5030 Accessed on: 01 Sep. 2022

20 – VEIGA, VIVIANE; DIB, SIMONE FAURY; LIMA, JEFFERSON; PENEDO, ERICK; HENNING, PATRÍCIA. PLANO DE GESTÃO DE DADOS FAIR DA FIOCRUZ: um desafio para a comunidade científica em saúde. PÁGINAS A&B ARQUIVOS E BIBLIOTECAS, v. 3, p. 249-250, 2020. Available at: http://aleph.letras.up.pt/index.php/paginasaeb/article/view/10172 Accessed on: 01 Sep. 2022

21 - HENNING, PATRÍCIA; SILVA, LUIS OLAVO BONINO DA; PIRES, LUÍS FERREIRA; SINDEREN, MARTEN VAN; MOREIRA, JOÃO LUÍS REBELO. The FAIRness of data management plans: an assessment of some European DMPs. RECIIS - Revista Eletrônica de Comunicação, Informação & Inovação em Saúde , v. 15, n. 3, p. 722-735, 2021. Available at: https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/2270/0 Accessed on: 01 Sep. 2022

CIÊNCIA E SAÚDE
PARA A SOBERANIA
E A DEMOCRACIA

Aqui somos SUS

ICICT Insttute of Communication and Scientific and Technological Information in Health

Pavilhão Haity Moussatché – Campus Fiocruz | Manguinhos - RJ
Av. Brasil, 4.365 – Rio de Janeiro – CEP: 21.045-360
Tel.: (55XX21) 3865-3131
Fax: (55XX21) 2290-1696

www.icict.fiocruz.br

ICICT
Insttute of Communication and Scientific and Technological Information in Health

Health Ministry
Fiocruz
Oswaldo Cruz Foundation

SUS