



This project has received funding from
The European Union's Horizon 2020 research and innovation programme under
Grant agreement No 101017562

Finding Endometriosis using Machine Learning

FEMaLe

Call/Topic: Digital transformation in Health and Care

Type of action: RIA

Date: 30-08-2022

Document information

DELIVERABLE NUMBER	10.30
DELIVERABLE TITLE	Data Management Plan: Initial
RESPONSIBLE AUTHOR	AU
GRANT AGREEMENT No.	101017562
DOCUMENT TYPE	ORDP
WORKPACKAGE N. TITLE	WP10 Governance: Management, Financial, Risk
LEAD CONTRACTOR	AU
AUTHOR(S)	AARHUS UNIVERSITY, CORRELATE AS
PLANNED DELIVERY DATE	May 31 st , 2022
ACTUAL DELIVERY DATE	August 31 st , 2022
DISSEMINATION LEVEL	Public
STATUS	First draft
VERSION	1.5
REVIEWED BY	Ulrik Kirk Bak, Emilia Alegria

Document history

Version	Date ¹	Comment	Author	Status ²
1.1	02-05-2022	First draft created	AU	Drafted
1.2	30-06-2022	Second draft created, including contribution from FEMaLe partners.	AU, CORR	Drafted
1.3	29-07-2022	Third draft prepared for FEMaLe Review Panel.	AU, CORR	Drafted
1.4	17-08-2022	Final draft created, based on FEMaLe Review Panel feedback.	AU, CORR	Completed
1.5	29-08-2022	Final version ready for submission, quality checked by FEMaLe PMO.	AU	Validated

¹ As per the project's cloud storage or per email date if applicable.

² Drafted, completed or validated as per the project's cloud storage or per email date if applicable.

Disclaimer

This document provides the description of the FEMaLe Initial Data Management Plan, DMP, which constitutes Deliverable D10.30. It was prepared, based on the template for the ERC Open Research DMP.³

The content of this deliverable does not reflect the official opinion of the European Union. Responsibility for the information and views expressed herein lies entirely with the author(s).

All FEMaLe consortium members are also committed to publish accurate and up to date information and take the greatest care to do so. However, the FEMaLe consortium members cannot accept liability for any inaccuracies or omissions, nor do they accept liability for any direct, indirect, special, consequential, or other losses or damages of any kind arising out of the use of this information.

The following sections will describe how we plan to make the project data Findable, Accessible, Interoperable and Reusable (FAIR). Each of the following five issues will be addressed with a level of detail appropriate to the project:

1. The DMP will elaborate how we will manage data life cycle within and outside the project.
2. The DMP will analyze relevant issues in terms of property rights, open access, and licenses.
3. The DMP will establish a shared approach to manage data.
4. The DMP includes forms and templates for proper data specification treatment.
5. The DMP is kept up to date during all the project implementations.

The use of the platform and the treatment of all data by the FEMaLe Consortium is regulated by the DMP, defined in coherence with the functionalities of the DMP.

The FEMaLe SharePoint site has been implemented, hosted by Aarhus University. The FEMaLe SharePoint server will be the primary site for keeping project documents, secure information, and enable exchange between FEMaLe Partners.

The main objective is to keep an updated filing system of all FEMaLe information, administrative activities, meetings, budgets, contractual documents, status of project and copies of all finalized deliverables, etc.

All FEMaLe partners can download files and documents, while the 10 WP leaders can upload documents on the FEMaLe SharePoint site, connected to the Correlate platform, currently a FEMaLe MVP, which means that new features and functionality will be implemented soon. This document will be updated accordingly.

Please note that the ERC Data Management Plan is not a part of the Ethics Review. It is the responsibility of the Principal Investigator to inform the ERCEA Ethics Team of any ethics issues/concerns regarding the collection, processing, sharing and storage of data in relation to the project. The Principal investigator can also be asked to submit an Ethics Data Management Plan (Ethics DMP).

³ Based on '[Guidelines on FAIR Data Management in H2020](#)', version 3.0. 26.07.2016, Annex 1.

TABLE OF CONTENTS

- 1. ACRONYMS AND ABBREVIATIONS 5
- 2. INTRODUCTION 5
- 3. DATA SUMMARY 6
 - 3.1. PURPOSE OF DATA COLLECTION AND RELATION TO PROJECT OBJECTIVES 6
 - 3.2 TYPES OF DATA 7
- 4. FAIR DATA 8
 - 4.1. MAKING DATA FAIR, INCLUDING PROVISIONS FOR METADATA 8
 - DATA-DRIVEN MANAGEMENT: AU SHAREPOINT SITE 8
 - 4.2 MAKING DATA OPENLY ACCESSIBLE 9
 - 4.3 MAKING DATA INTEROPERABLE 11
 - 4.4. INCREASE DATA RE-USE 11
- 5. ALLOCATION OF RESOURCES AND DATA SECURITY 12
- 6. ETHICAL ASPECTS 12
- 7. DATA PROVIDE FOR EACH WP 12
- ANNEXS 14

1. Acronyms and abbreviations

Abbreviation	Description
EC	European Commission
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable and Re-usable
ICT	Information and Communications Technology
XML	Extensible Markup Language
CSV	Comma-separated values
FNC	File Naming Convention
DOI	Digital Object Identifier
IP	Intellectual Property
PC	Project Coordinator
GDPR	General Data Protection Regulation

2. Introduction

FEMaLe project complies with the FAIR data management concept to develop this DMP. FAIR data management requires the project data to be: Findable, Accessible, Interoperable and Re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution.

FEMaLe deliverable 10.30 is not intended as a strict technical implementation of the FAIR principles, it is inspired by FAIR as a general concept. The following documents have been referred to develop the FEMaLe DMP:

1. Guidelines on FAIR Data Management in Horizon 2020 (EC, 2016).
2. FAIR data principles (FORCE11, no date).
3. FAIR principles (Wilkinson et al., 2016).
4. ERC Open Research Data Management Plan (DMP) template (ERC, 2017)

The ERC Open Research Data Management Plan (DMP) template (ERC, 2017), which was used as template for this document, contains a set of questions that should be answered on every project using FAIR data management with a level of detail appropriate to the project.

The DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, DMPs should have a clear version number and include a timetable for updates. As a minimum, the DMP should be updated in the context of the periodic evaluation/assessment of the project. If there are no other periodic reviews envisaged within the grant agreement, an update needs to be made in time for the final review at the latest.

3. Data Summary

3.1. Purpose of data collection and relation to project objectives

FEMaLe will generate a broad spectrum of data from numerous mainly public sources and will develop these into database structure or guidelines. FEMaLe will also collect numerous stakeholder data including interview information. The initial DMP for the FEMaLe project will include:

- The data to be collected during the FEMaLe project, including the type and format of the data and its association with the project's programme of work.
- An outline of how the data will be made findable, including provision for metadata.
- Made openly accessible.
- Made interoperable (including specifying what data and metadata vocabularies, standards, naming conventions or methodologies followed).
- Rules for data re-use, exploitation and sharing, indicating whether the data produced and/or used in the project is useable by third parties, when data will be made available and restraint periods where required.
- The allocated resources associated with making the FEMaLe data FAIR and the beneficial responsible for data management within the project.
- Data storage, recovery, and security measures for preservation.
- Management of ethical issues related to data preservation, protection and sharing.

The DMP will outline the consortium's data management policy for the newly generated datasets, enabling continuous access to quality-assured data from the FEMaLe project.

The FEMaLe project involves ICT (Information and Communications Technology) solutions, using a variety of technologies. CORRELATE AS offers a tailored software as a service (SaaS), giving the users an instant transparent and augmented layer over other online and actionable content, files, and pages, stored in and across unconnected and scattered private and public online platforms.

The importance of making gathered datasets FAIR (Findable, Accessible, Inoperable and Reusable) is recognized and planned for, but considering the demands of the living nature of a Data Management Plan as the FEMaLe project progresses, the final form of the datasets cannot yet be declared in detail. The Project Coordinator (PC) will be able to offer an outline of the data once it has been collected.

To date, there has been no data gathered. Consequently, none have been reported in the project deliverables. This deliverable and its associated data will contribute to the further project activities of the relevant WPs.

Publicly available information about the project is made available via the FEMaLe Project Website, available at: <https://findingendometriosis.eu>

The data repository will be FEMaLe's SharePoint site at Aarhus University in which all the documents will be archived and kept. The Correlate FEMaLe platform will be the place in which the Consortium members can share, manage, and operationalize knowledge in-between them, and it will be linked to the relevant folders in FEMaLe SharePoint to facilitate the communication online.

3.2 Types of Data

FEMaLe will make use of and generate a large variety of data. We will use the online tool www.dmponline.dcc.ac.uk to generate the initial DMP, using the H2020 DMP template.

The project will generate and collect a diverse array of types of data, including media content, statistical data, documents, reports, etc., as well as several types of formats, including .xls, .doc, .shp, .csv, .dta, .mp3, etc. At this point of the project, we are unable to provide specific example datasets. We are in the process of defining the data that will be generated and collected. Specifics will be clear at a future date.

A first overview of the data to be used and generated in FEMaLe is shown in table 1.

Work Package	Data Types	Suitable for Open Data	Data Format	Size of Needed Storing Capacity
WP1				
WP2				
WP3				
WP4				
WP5				
WP6				
WP7				
WP8				
WP9				
WP10				

Table1: Data to be generated in the WPs – first overview.

4. FAIR DATA

4.1. Making data FAIR, including provisions for metadata

FEMaLe data management aims to comply with the FAIR principles and make research data findable, accessible, interoperable, and re-usable. Presently, FEMaLe has opted into the open research data pilot (ORDP) and aims to make data open whenever possible, but as closed as necessary when taking into consideration personal data and privacy. If it affects personal confidentiality and privacy, data will not be shared publicly.

The project will be collecting data from different sources. These can include news, articles, statistical data from governmental organizations in various regions, etc. Specific sources will be specified at a later point of the project. There is no established mechanism to catalog the information produced in a common way (there is no task whatsoever intended for this procedure). However, it can be said that the data generators within the project (living labs and policy labs) are aware of the need to consider publishing their data following the maxims of persistence and unique identification.

The ISO / TC211 specifications for the provision of metadata for information in the areas of geomatic and digital geographic information will be considered. At the European level, these regulations will be followed:

- *European Commission (EC) Regulation No. 976/2009 of October 19, 2009, implementing Directive 2007/2/EC of the European Parliament and of the Council regarding network services.*
- *European Commission (EC) Regulation No. 1089/2010 OF THE COMMISSION of November 23, 2010, applying Directive 2007/2 / EC of the European Parliament and of the Council as regards the interoperability of spatial data sets and services.*
- *European Commission (EC) Regulation No 1205/2008 of December 3, 2008, implementing Directive 2007/2 / EC of the European Parliament and of the Council regarding metadata.*

Document file names will be kept short to avoid unnecessarily long paths and always include the last person to edit the document.

Data-Driven Management: AU SharePoint site linked to the FEMaLe Correlate Platform

Some of the project documents from FEMaLe's SharePoint site at Aarhus University will be linked to the FEMaLe Correlate platform, making files easily accessible and findable and aid in the co-operation and co-creation of projects and workflows. Correlate provides a secure, easy, and flexible solution to synthesize data and information through a non-intrusive, lightweight, high performance, cost-efficient, easy-to-use desktop and mobile friendly web app that connects multi-cloud services.

Correlate.com is a Software as a Service (SaaS) application enabling a transparent and augmented layer on top of other online digital activities, systems, and apps. The user correlates their digital asset with a link, which is dragged and dropped into private or team-shared boards where information is further organized, annotated, and accessed by all collaborators. WP leaders also started creating their own Teams and Boards according to their specific needs and have directly invited the members of other partner organizations they are cooperating with.

4.2 Making Data Openly Accessible

FEMaLe will, as far as possibly, make produced data openly available. All data generated by the project will be accessible to partners after quality control according to the Grant agreement. Data, apart from that which includes sensitive data and general personal data affected by GDPR, or raising any ethical concerns, will be shared with relevant partners and stakeholders.

The project's data will be deposited in the AU repository as in relevant file formats (e.g. Excel / CSV) at FEMaLe's SharePoint server. This is in accordance with the GDPR-compliant regulations of the host and in accordance with the Aarhus University regulation on data management. The specific repositories to store the data and associated metadata, documentation and code is yet to be defined. At this stage of the project the deliverables are stored in AU SharePoint site, the primary repository.

In accordance with the Grant Agreement all research related data, excluding personal and sensitive data, will be stored for at least five years after the end of the research project (in case there is a high interest in the datasets or due to different national legislation, data may be stored for a longer period, which will be transparently discussed and approved within the consortium and relevant parties). Data that are used for publication will be stored at least five years after publication.

Open science is the movement to make scientific research, data, and dissemination accessible to all levels of an inquiring society, amateur or professional. Open science is transparent and accessible knowledge that is shared and developed through collaborative networks.

Open access refers to research outputs which are distributed online and free of cost or other barriers, and possibly with the addition of a Creative Commons license to promote reuse. Open access can be applied to all forms of published research output, including peer-reviewed and non-peer-reviewed academic journal articles, conference papers, theses, book chapters, and monographs.

For data collected in work packages will be required to remain closed, for example where an interview transcript cannot be sufficiently anonymised, and according to national regulation.

Open access to publications: Any proposed publication or communication by one of the Beneficiaries must be submitted to other beneficiaries for their consent, according to FEMaLe CA Article 8 Dissemination of own results. All publications will be either gold or green open access in accordance with the H2020 requirements.

To encourage re-use and further application of project results, the guideline for FEMaLe Open Data is as follows:

a. All data used for scientific publications based on FEMaLe activities will be made available via open-access online platforms. Exceptions will, however, occur for restricted data, including: i) lack of possibility to anonymise the data ii) due to the protection of intellectual property rights following the consortium agreement, e.g. data related to new products or specific business models; iii) due to institutional regulation, or iv) if release of all or part of the data to open-access platforms would obstruct later project tasks.

b. Data that results from FEMaLe activities and that underlies scientific publications must be submitted to the relevant repository no more than 60 days following any related publication in scientific journals unless the beneficiaries have outlined justifiable reasons for maintaining data confidentiality. Such reason could for example be that a dataset is subject to protection, due to lack of possibility to protect personal data, or that some data are expected to result in multiple publications. In the latter case, the ultimate embargo period will be 60 days after the last publication deriving from the project. The WP-leader is responsible for the dataset being stored, while the data producer is responsible for conducting the upload.

c. For datasets that are stored at the FEMaLe SharePoint for internal data sharing, we will use a naming standard that makes it easy to identify the dataset. This will follow the format: **WPNumber_TaskNumber__PartnerName_DatasetName_Version__DateOfStorage**, where the project name is FEMaLe, the Partner Name represents the name of the data custodian: WP Leader or Task Leader. An example of this naming format would be: WP1_T1.1_AU_task report _v1_01.02.2022 (date).

d. A beneficiary that intends to disseminate their results must give advance notice to the other FEMaLe Beneficiaries and the Coordinator of – unless agreed otherwise - at least 30 days, together with sufficient information on the results it will disseminate. Any other beneficiary may object within — unless agreed otherwise — 20 days of receiving notification if it can show that its legitimate interests in relation to the results or background would be significantly harmed.

e. In such cases, the dissemination may not take place unless appropriate steps are taken to safeguard these legitimate interests; all as decided in the Consortium Agreement article 8 Dissemination and following the Grant Agreement article 29.1.

f. Beneficiaries who intend to protect their data should notify all Consortium Beneficiaries, the Project Coordinator, and the Executive Board as soon as possible to ensure that the optimum level of confidentiality is upheld from an early stage.

4.3 Making Data Interoperable

FEMaLe will allow data exchange and re-use between researchers, institutions, organisations, countries according to the Grant agreement and the Consortium agreement. It means FEMaLe is adhering to general standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins.

The data and metadata, standards, and methodologies the FEMaLe project follows are with the aim to make the project data interoperable and standard vocabularies for all data types present in the project will allow inter-disciplinary and trans-disciplinary interoperability.

Data produced through interviews, user/focus groups will use standard formats (e.g. TXT, DOC, XLSX JPEG, RAW, AVI formats) and made compliant with available (open), software applications, facilitating the recombination with various datasets from different origins. The data produced will be stored on FEMaLe SP.

4.4. Increase data re-use

Whenever suitable, data will be Open Access licensed data after considering of personal data, intellectual property rights and any additional legal and ethical requirements. Based on Open Access regulation, we aim to allow data to be re-used by third parties, but with restrictions if IPR or other rights demanding such restrictions.

Data licensing will be based on guidance provided by Aarhus University. Copyright of the data are based on EU H2020 guidelines and Digital Curation Centre (DCC), an internationally-recognised centre of expertise in digital curation with a focus on building capability and skills for research data management.

Access to the research data will be dependent on any agreed 'embargo period' based on national and EU regulations. The 'embargo period' is applied to give time to publish the work or seek patents, where applicable and this will be as short as possible until the work is accepted for publication or patent, bearing in mind that research data should be made available as soon as possible.

Data will be stored based on the contractual terms, until which it can be re-usable (all research related data will be stored at least for five years after the end of the research project).

In case of sharing data or restricting certain data with third parties outside of the consortium, a data sharing agreement will be set up that will detail anonymizing or aggregating data, participant consent for data sharing, copyright permissions, and agreement on an embargo period. Data will be used in standard forms allowing reuse, as well as allowing searchability.

Each WP or task leader is responsible for preparing data generated in the project in a form that follows agreed guideline for storage and reuse of the data

5. Allocation of resources and Data security

The Consortium will use the internal repository for making the dataset accessible. Data will be stored in FEMaLe SharePoint server, hosted by Aarhus University. All data will be stored on a test and a production server. The server is regularly backed up via standard backup procedures by the AU IT department, which ensure that the data are safely stored.

As for the publications, where the analyses of the empirical research data will be presented, the FEMaLe consortium will publish them in scientific journals that allow open access the costs related to open access will be claimed as part of the H2020 grant.

6. Ethical aspects

The present chapter intends to provide and answer to a number of questions as:

1. Are there any ethical or legal issues that can have an impact on data sharing?
2. Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

In this framework, there are no ethical issues but informed consent for data sharing will be collected if personal data will be needed from stakeholders.

7. Data provide for each WP

To build the Initial DMP, AU will distribute a data inventory survey to all relevant FEMaLe partners to obtain comprehensive and up-to-date information on data planned to be produced in FEMaLe. Work package (WP) leaders is the first point of contact for WP10, to enable efficient ongoing (i.e., for the duration of the project) agreement and liaison on FEMaLe data collection, management, and related information or issues.

AU will contact the WPs with information collection templates, instructions for completion, and links to further online guidance on data collection and the use and development of experimental data templates.

These data inventory survey materials (Annex 1) will be distributed to the Data Management Responsible (DMR), appointed within each WP. The DMR will help and coordinate data gathering and collation for their WP, returning the information on completion of the exercise to WP10 DMP.

The templates will be used to obtain information from partners on the experiments and test planned the types and formats of data they plan to generate with outlines timescales, where available. Besides providing the initial information for DMP, the templates that aims to enhance and incorporate appropriate naming conventions, metadata, and open data features.

The information gathered will be collated, and extracts for each WP are summarized in the Initial DMP (see Annex 2).

At this stage in the project, information in some areas is incomplete or unconfirmed, but this will be supplemented as work proceeds. For example, partners do not yet know some data volumes or formats; some WPs or partners are still developing or agreeing upon their detailed experimental programme; or because a plan will be fully determined only after the results of other initial investigations are available. However, this first version forms a very substantial basis for the DMP at this stage and the information will be supplemented as liaison continues between Partners and WP10-AU as the FEMaLe project proceeds, with periodic updates and amendments to the DMP.

The DMP will be a live document reviewed periodically and updated as necessary.

ANNEXS

ANNEX 1

FEMaLe WP10 - Data Inventory Survey and Data Collection Template Developments

Dear [WP leader]

On behalf of FEMaLe WP10, I am writing to you to appoint a Data Management Representative (DMR) for your WP.

We must produce the mandatory Data Management Plan (DMP), Deliverable D10.30. To do this, we need to identify existing and new requirements for data collection templates, SOPs, and, possibly, Curation System, etc. This to plan and schedule the FEMaLe storage. We need your and the DMRs help to identify the types of data analysis and tools of use in FEMaLe. This to obtain correct input and overview of data at the WP level and good coordination and data curation for each WP.

We have designed forms for data collection at various levels:

1. An overview at WP level by the Data Management Representative of the WP's overall experimental work programme, with a summary of all the data collection exercises being conducted.
2. Details from the beneficiaries collecting/producing data in WPs

Could you please assist us with distributing form Bs (with instruction sheet) to the partners that collect data, and, after this, collect completed forms from your beneficiaries? Further, please ensure that the DMR fill in form A for your WP and return it as soon as possible.

It is not expected that this will take long to do this, the forms are brief, and it should take the beneficiaries just a few minutes to complete for each type of experiment.

Instructions for completion are outlined in the document FEMaLe-DMP introduction.doc and in Forms A and B. These documents are also available at the FEMaLe server.

The information collection will take place over the next weeks for completion by 10.09.2022 and will be inserted in the DMP deliverable10.30. It is understandable that it may not be possible to know the kind and amount of data at present, in this case please return the forms when this is known. In any case, we will get back to you later to consider the overall WP perspective, an addition to the form and timetable for the Data Management Plan as the project progresses.

Please proceed to contact DMR and the beneficiaries with the forms and let us know if you need help completing the forms.

Best regards,

DATA INVENTORY OF PROJECT TEST RESULTS & DATASETS

FEMaLe Introduction to Data Inventory Survey of Project Results & Datasets, and Form A and B completion instructions.

FEMaLe is creating a project database that will make its data available initially to the Consortium and later to other interested parties following also GDPR rules. To facilitate this, we are making a catalogue of the data collected or produced. This will enable us to:

- a. Develop the Data Management Plan (DMP).
- b. Identify existing and new needs for data collection templates.
- c. Obtain and curate data and upload them to the database as the project progresses.
- d. To identify the various types of FEMaLe data analysis and tools (e.g., guidelines, interviews etc).

It is very important that we obtain this information from the WP beneficiaries

Each WP Partner carrying out data generation/gathering must contribute to the inventory, so we build a complete picture for the Project.

Each WP leader must assign a Data Management Representative (DMR) that have the knowledge of work programme for this WP. This to help coordinate with the collection of information from the beneficiaries. The DMR should provide overall information about the WP's programme, timetable, and data.

This will allow FEMaLe to link with data collection template requirements, so we can develop improved and new templates where required. This will enable a much better project data storage.

It is not envisaged be to a lot of work: Form B is brief, so it will only take the beneficiaries no more than a few minutes to complete.

The forms A and B should be returned to at latest by 10.09.2022.

You may also find the Inventory Forms and Instructions at the FEMaLe serve.

Kind regards and thank you

FORM 1A

Work Package Level Information	
Work package (Name, Number, Leader).	
Nominated Data Management Representative (Name and e-mail).	
Completion of form date (Day-Month-Year).	
A short overview of the WPs programme – in relation to data production/gathering.	
A short summary of possible experiments that will be performed.	

FORM 1B

#	WP	Beneficiary No.	Data type	Description & Purpose	Date of initial production /end of production	Data format	Expected number of data	Restriction foreseen	Reason	Alternatives	Data location	Accessibility	Mean of access	Metadata
1														
2														
3														
4														
5														
6														
7														
8														

FORM 2

	Yes/No	Give as detailed information as possible
WP		
Beneficiary ID		
Contact person (Name, e-mail, phone, address)		
Name of data gathering		
Date of initial production		
Date of final production		
Main type data being produced		
Guideline or SOPs being used		
Collection and storage procedure of the data		
Any specific template used?		

Data summary

WP	Data Type	Origin	Format
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

FAIR DATA

WP	Data type	Open	Reason	Alternative
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				