Acronym:	COTIDIANA
Name:	Mobile Patient-centred System to Improve Drug Trials and Care of Older-adults with Rheumatic Diseases
Call:	AAL Call 2020
Contract nr:	AAL-2020-7-146-CP
Start date:	01 April 2021
Duration:	28 months

D1.6 Data Management Plan (a)

Nature[1]: P
Dissemination level[2]: PU
Due date: Month 6
Date of delivery: Month 11
Partners involved (leader in bold): AICOS, Definition12, Raffeiner, MUW, **Pryv**, UNL-NMS-CHRC
Authors: Francisco Nunes (FhP), Evelina Georgieva (Pryv)

[1] L = Legal agreement, O = Other, P = Plan, PR = Prototype, R = Report, U = User scenario

[2] PU = Public, PP = Restricted to other programme participants (including the Commission Services), RE = Restricted to a group specified by the consortium (including the Commission Services), CO = Confidential, only for members of the consortium (including the Commission Services)

Partner list

Nr.	Partner name	Short name	Org. type	Country
1	Associação Fraunhofer Portugal Research	AICOS	Research	Portugal
2	Definition 12 AG	Definition12	SME	Switzerland
3	Mag. Andreas Raffeiner	Raffeiner	SME	Austria
4	Medizinische Universität Wien	MUW	University	Austria
5	Pryv SA	Pryv	SME	Switzerland
6	Universidade NOVA de Lisboa	UNL-NMS-CHRC	User	Portugal

Revision history

Rev.	Date	Partner	Description	Name
1	28.02.2022	Pryv	Complete version	Evelina Georgieva

Disclaimer

The information in this document is subject to change without notice. Company or product names mentioned in this document may be trademarks or registered trademarks of their respective companies.

Glossary

DMP: Definition of Data Management Plan

Table of contents

1. Summary 6
2. Consortium Data 7
3. Ensuring privacy and data management compliance awareness and excellence 8
4. Data management Plan 10
4.1 Project Documentation 10
4.2 Laboratory data collection dataset 13
4.3 User research with end-users 17

1. Summary

The following DMP describes the data management lifecycle for all data sets that will be collected, processed or generated during the COTIDIANA project "**Mobile Patient-centered System to Improve Drug Trials and Care of Older-adults with Rheumatic Diseases**".

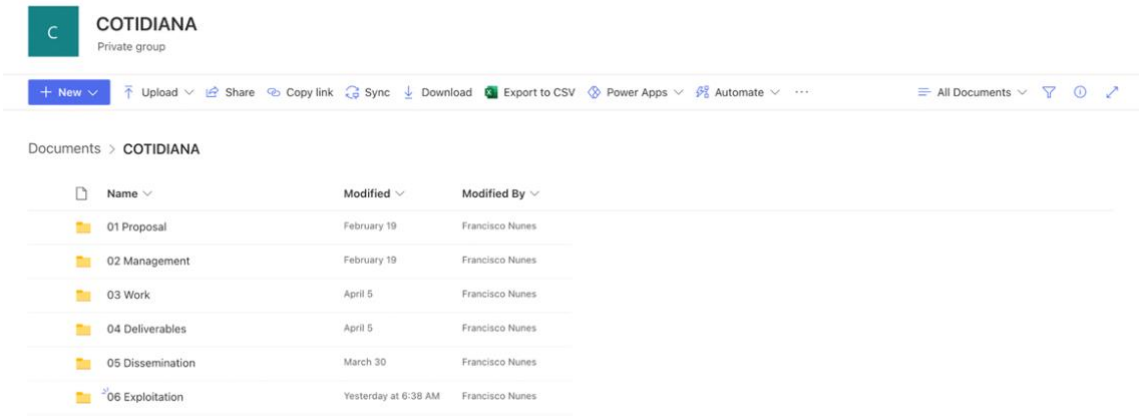
The DMP provides a general framework and description of the data management, regarding consortium data management, used and generated research data, that will be applied during the COTIDIANA project including: Data management and strategy, data collection, data storage, data processing; sharing, quality control, privacy compliance and governance.

The DMP is an evolving document, therefore, some aspects may be added and/or updated in later versions of the document.

The DMP includes an overview of the data governance throughout its entire lifecycle and reviewed on the basis of per dataset specifications.

2. Consortium Data

We have created a Collaborative Sharepoint digital workplace, using Fraunhofer's infrastructure to ensure all partners access to real-time updated documentation. Appropriative security measures are put in place: the access is password protected, the data is encrypted, data backups.



The data formats used are:
-.docx
-.xlsx
-.PDF
-.pptx

3. Ensuring privacy and data management compliance awareness and excellence

Ethical & Privacy Board:
We set-up an internal ethics and privacy board, composed of participants from each partner of the consortium to discuss ethical issues as they appear.

Workshops & Awareness Training:
We have already held a series of ethical and privacy discussions, in topics such as GDPR and Privacy and continue elaborating on creation of ethics training materials. Topics of informed consent, personal data collection, and data management have been discussed, and materials were made available to partners.

Applicable Data Management & Privacy Regulations:
We have created a shared file to identify, classify and set relevant regulations/standards/guidelines/practices of the EU, H2020, Grant Agreement in Switzerland, Austria & Portugal.

We collected and investigated all national and internal policies each organization is under so that we can have a detailed overview of any question of patient's consent, data collection, sharing and storage which could affect the product development and data lifecycle. The document includes the information regarding:

- Procedures compliant with EU and local Laws
- Context COTIDIANA
- Involvement in the Clinical Trial COTIDIANA testing planning
- Analysis on the collected and governed data which includes:
  - Deliverables and presentations
  - Interview transcripts, survey responses and similar
  - Laboratory data collection dataset – different types of data
  - Data collected during the field trials- different types of data
- Practice for pseudonymization and anonymization (All data we are collecting is pseudoanonymised. The survey we will distribute is anonymised, but since we also gather some personal data it can also be considered pseudoanonymised)
- Detailed DMP Q&A per dataset, following the process of EU validated Argos tool

A detailed DMP document has been created to serve the Consortium:

- as a live document to advise with at any given time when a question concerning data privacy arises
- to identify the requirements for the system architecture development (in particular the data-gathering platform)
- specify to what extent and under what conditions and through what mechanisms data once generated in the project COTIDIANA can be (re)used/shared within the project/consortium (e.g. how can data generated in

one use case be re-used in a second use case within the project for analytics purposes) and beyond (e.g. for publications and reporting
- be in tune with the testing and demos planned in COTIDIANA
- to summarize the relevant information and present it in a easy-to-use table-format that can be used by the partners to ensure that the implement necessary technical and organizational measures in the context of data handled and shall serve the consortium as a general manual/guideline and shall provide flexibility to be adapted with the advance of the project

4. Data management Plan

The Review of the Data Management Plan is currently segregated into data categories: Project Documentation, Laboratory data collection dataset and User Research. We foresee that with the evolution of the project, new data categories will be added.

The methodology for describing the lifecycle of each dataset is based on using Argos tool

4.1 Project Documentation

This dataset consists of the documentation created during the COTIDIANA project. It includes the internal documentation, created for the purpose of sharing initial results or discussing project-related aspects. Moreover, it also includes the official documentation created during the project such as the deliverables.

Data Summary

1.1.1 What is the purpose of the data collection/generation and its relation to the objectives of the project?

The data collection/generation has three purposes. The internal documentation is meant to support discussions about the project, initial results, or keeping record of insights and decisions. The external documentation, materialized in the deliverables, is meant to be shared with the general public and with the AAL Central Management Unit that oversees the execution and evaluates the impact of the project.

1.1.2 What types of data will the project generate/collect? [observational (e.g., sensor data, data from surveys), Other Reports, meeting notes, and deliverables

Observational data consists of excerpts from the data collected during the project for discussion (e.g., interview quotes, tables with data from laboratory data collections, summary of results from field trials).

1.1.3 What formats of data will the project generate/collect? [Text files, Numerical, Multimedia

Research protocols, instructions, reports, and deliverables will be stored as Word (.docx), Latex (.tex, .bib), or Adobe Portable Document Format (.pdf). User interface screenshots or pictures will be stored in JPEG (.jpg), or PNG (.png).

1.1.4 What is the origin of the data?

Secondary data

1.1.5 What is the expected size of the data?

MB (megabyte) Comment: (1) Research protocols, instructions, reports, and deliverables: each file will amount to 500KB, up to a total of 50MB; (2) User interface screenshots or pictures: each file will amount to 5MB, up to a total of 200MB;

1.1.6 To whom might it be useful ('data utility')? [Researchers, The public, Industry]

The protocols, instructions, reports, or deliverables will be useful for researchers or industry innovators working on similar contexts to COTIDIANA. The public might also be interested in the results of the project, and thus to learn from the documentation created during the project.

2.1 Reused Data

2.1.1 Will you re-use any existing data and how?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 Will you use metadata to describe the data?

No

3.1.1.3 Will your metadata use standardised vocabularies?

No

3.1.1.7 Will you use naming conventions for your data?

Yes

3.1.1.8 Please provide more details and examples on used naming conversions

[Number of workpackage].[Number of deliverable of workpackage]_[Name of deliverable]_[version].PDF

3.1.1.9 Will you provide clear version numbers for your data?

Yes

Comment: Versioning is included in the naming convention of the files created.

3.1.1.10 Will you provide persistent identifiers for your data?

No

3.1.1.12 Will you provide searchable metadata for your data?

No*Comment: This is not likely to be required considering the simple nature of the documents to be shared (i.e., project deliverables).*

3.1.1.15 Will you use standardised formats for your data?

No

3.1.1.17 Please describe the formats you plan to store your data in, including any URLs to documentation.

Commonly used formats

3.1.1.18 Are the file formats you will use open?

Yes

3.1.1.20 Do supported open-source tools exist for accessing the data?

Yes

3.1.1.22 Will you provide metadata describing the quality of the data?

No. *Comment: Not applicable.*

3.1.2 Making data openly accessible

3.1.2.1 Are there ethical or legal issues that can impact sharing the data?

Yes

Comment: Some of the documentation created in the project is confidential and thus it is not going to be shared.

3.1.2.2 Will your data be openly accessible?

some

Comment: Public reports or deliverables

3.1.2.3 Please provide the URL of the data which can be made available

https://cotidiana.eu/results.html

3.1.2.4 How will the data be made available?

[Project website, Other]
http://www.aal-europe.eu/public-deliverables/

3.1.2.6 Is the storage sufficiently secure for the data and does the storage provide backup and recovery procedures?

secure with backup and recovery

3.1.2.7 Are there any methods or tools required to access the data?

No

3.1.2.10 Will you also make auxiliary data that may be of interest to researchers available?

no auxiliary data

3.1.3 Making data interoperable

3.1.3.1 Will you use a controlled vocabulary for your data?

No

3.1.3.2 Will you provide a mapping to more commonly used ontologies?

No

3.1.4 Increase data reuse

3.1.4.1 When do you plan to make your data available for reuse?

after article publication

3.1.4.5 Do you have documented procedures for quality assurance of your data?

No. Dataset is too simple to require this

3.1.4.8 Will you provide any support for data reuse?

No

4.1 Allocation of resources

4.1.1 How will the cost of making your data findable, accessible, interoperable and reusable be covered?

The sharing of deliverables on the website of the european funding mechanism will ensure there are no costs and information is shared during a long period.

4.1.2 Will you identify a data manager to manage your data, if not who will be responsible for the management of your data?

Yes. *Comment: Project coordinator*

4.1.4 How do you intend to ensure data reuse after your project finishes?

The sharing of deliverables on the website of the european funding mechanism will ensure there are no costs and information is shared during a long period.

5.1 Data Security

5.1.1 What do you plan to do with research data of limited use?

Kept on secure, managed storage for limited time

6.1.1 Are there any ethical or legal issues that can have an impact on data sharing?

Yes. *Comment: Confidentiality of the documents.*

7.1 Other 7.1.1 Do you make use of other procedures for data management? No

4.2 Laboratory data collection dataset

This dataset is a result of a laboratory data collection conducted with participants living with Osteoarthritis (OA), Rheumatoid Arthritis (RA), and Psoriatic Arthritis (PsA), performing a set of tasks. The tasks included (1) physical activities, such as walking for some meters with a smartphone in the pocket, (2) hand dexterity exercises, such as copying some text sentences with the smartphone keyboard, or (3) answering Patient Reported Outcome Measures in a smartphone app. Some participants also volunteered to donate processed and anonymized communication logs from their smartphones. In addition to these, participants filled-in a set of validated scales to gather data about participant's current health status. Collected data provides a rich characterisation of a large group of patients with these conditions, and will support the future development of algorithms that help understand the condition state of patients with OA/RA/PsA, only using the smartphone.

1.1 Data Summary

1.1.1 What is the purpose of the data collection/generation and its relation to the objectives of the project?

[To obtain information, To share information, To develop a product]

Comment: The purpose of data collection had three objectives. Since we had the goal of creating passive sensing algorithms for the smartphone, it was important to collect data from participants that we could explore and attempt correlations with their current health state. Clinicians in the project recognised that this dataset could advance the understanding of patients' conditions, and that the dataset could be used for clinical research, so we embraced that goal. Moreover, we believed in the potential of open science to build better solutions in the future, so we created the dataset with the goal of sharing the data as well.

1.1.2 What types of data will the project generate/collect?

[observational (e.g., sensor data, data from surveys), derived or compiled (e.g., text mining, 3D models)]

Observational data generated consists of: answers to scales, data from inertial sensors, logging data from the keyboard, and observation annotations. We will also record video footage of participants performing physical activities, which will be deleted after annotations are made. Derived or compiled data generated consists of: processed communication logs, processed app usage logs, and processed gait and physical activity metrics.

1.1.3 What formats of data will the project generate/collect?

[Text files, Numerical, Multimedia]

Research protocols, instructions, and reports will be stored as Word (.docx), Latex (.tex, .bib), or Adobe Portable Document Format (.pdf). Raw data collected will be stored as Comma-separated values (.csv), JSON files (.json). Processed data will be stored as Comma-separated values (.csv), JSON files (.json), or SPSS (.sav).

1.1.4 What is the origin of the data?

[Primary data]

1.1.5 What is the expected size of the data?

MB (megabyte)

Comment: (1) Answers to scales: each file will amount to 100KB, up to a total of 10MB; (2) Inertial sensor data: each file will be amount to 5MB, up to a total of 500MB; (3) Logging data from the keyboard: each file will amount to 100KB, up to a total of 10MB; (4) Observation annotations: each file will amount to 100KB, up to a total of 10MB; (5) Processed app usage logs: each file will

amount to 100KB, up to a total of 10MB; (6) Processed gait and physical activity metric: each file will amount to 100KB, up to a total of 10MB.

1.1.6 To whom might it be useful ('data utility')?

[Research communities]

The dataset might be useful for two research communities. (Bio)medical informatics researchers, can be interested in the dataset, as it can enable them to explore correlations between validated scales and sensed data. Clinical researchers might be interested in the data as it can help better understanding rheumatic conditions.

2.1 Reused Data

2.1.1 Will you re-use any existing data and how?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 Will you use metadata to describe the data?

No

All data will be uploaded together with the relating metadata, including project context and best practices used.

3.1.1.3 Will your metadata use standardised vocabularies?

No

We did not find field-specific metadata vocabularies at the time of creating this Data Management Plan. In any case, the metadata accompanying the dataset will include: Title, Description and data explanation, Authors, Contact information, Publication date, Version, Grant acknowledgement, and related publications (if applicable).

3.1.1.7 Will you use naming conventions for your data?

Yes

3.1.1.8 Please provide more details and examples on used naming conversions

We believe the data will be easy to recognise data created via the following naming convention:
P[ParticipantNumber]_[Gait|PhysicalActivity|Keyboard|CommunicationLog|AppUsage]_[Raw Data|Processed|Analysed]_[VersionNumber].[FileExtension].

3.1.1.9 Will you provide clear version numbers for your data?

Yes*Comment: Versioning is included in the naming convention of the files created.*

3.1.1.10 Will you provide persistent identifiers for your data?

Yes

Comment: We are expecting to release the dataset with its own DOI number. Parts of the dataset will be shared in research publications, which will have a DOI number as well.

3.1.1.11 Persistent identifiers

DOI

3.1.1.12 Will you provide searchable metadata for your data?

No

Comment: This is not likely to be required considering the simple nature of the dataset.

3.1.1.15 Will you use standardised formats for your data?

No

3.1.1.17 Please describe the formats you plan to store your data in, including any URLs to documentation.

Commonly used format

3.1.1.18 Are the file formats you will use open?

Yes

Comment: The data produced will be stored and shared in open file formats.

3.1.1.20 Do supported open-source tools exist for accessing the data?

Yes

3.1.1.22 Will you provide metadata describing the quality of the data?

No

3.1.2 Making data openly accessible

3.1.2.1 Are there ethical or legal issues that can impact sharing the data?

Yes

Comment: The dataset consists of pseudo-anonymized data from a group of patients. Ethics review board has been sought and Informed Consent obtained from all participants. To minimize data sharing issues, data collection was planned so that it would reduce the theoretical possibility of participants being de-anonymized, by collecting aggregated data variables instead of identifying personal ones (e.g., age group vs date of birth). Still, when all data is obtained we will consider the ethical issues of sharing the data for all involved stakeholders.

3.1.2.2 Will your data be openly accessible?

Some *Comment: The collected data may be additionally processed to protect the privacy of participants, but, as much as possible, we will try to share all the collected data.*

3.1.2.4 How will the data be made available?

[Project website, Repository of Archive]

3.1.2.6 Is the storage sufficiently secure for the data and does the storage provide backup and recovery procedures?

secure with backup and recovery

3.1.2.7 Are there any methods or tools required to access the data?

No

3.1.2.10 Will you also make auxiliary data that may be of interest to researchers available?

no auxiliary data

Comment: Does not seem to be relevant for the dataset collected.

3.1.3 Making data interoperable

3.1.3.1 Will you use a controlled vocabulary for your data?

No

3.1.3.2 Will you provide a mapping to more commonly used ontologies?

No

3.1.4 Increase data reuse

3.1.4.1 When do you plan to make your data available for reuse?

after article publication

3.1.4.4 What internationally recognised licence will you use for your data?

BSD Zero Clause License

3.1.4.5 Do you have documented procedures for quality assurance of your data?

No*Comment: Dataset is too simple to require this.*

3.1.4.8 Will you provide any support for data reuse?

No

4.1 Allocation of resources

4.1.1 How will the cost of making your data findable, accessible, interoperable and reusable be covered?

[Other]

Comment: We will use tools that enable free sharing of our dataset, so that costs are kept small. Questions about the dataset and future support will be ensured by other research grants.

4.1.2 Will you identify a data manager to manage your data, if not who will be responsible for the management of your data?

Yes

Comment: Team member responsible for data collection.

4.1.4 How do you intend to ensure data reuse after your project finishes?

[Other, Data Center Archive Storage]

Comment: Zenodo & Paper publisher data repositories

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on data sharing?

Yes

Comment: See above.

6.1.2 What are the methods used for processing sensitive/personal data?

[Anonymising data where necessary]

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

4.3 User research with end-users

This dataset corresponds to the data collected during the ethnographic and participatory research study involving patients, carers, clinicians, clinical researchers, and Pharma which is part of project COTIDIANA. To inform the design of the system, we conducted qualitative interviews with: 1) patients with rheumatic conditions in Portugal, 2) clinicians involved in rheumatology care, and 3) stakeholders from hospitals, clinics, clinical trial units, Pharma, and CROs, who are involved in clinical research or drug trials. A survey was also distributed to the larger patient community in both Portugal and Austria, to check some of the collected results.

1.1 Data Summary

1.1.1 What is the purpose of the data collection/generation and its relation to the objectives of the project?

[To obtain information, To keep on record, To make informed decisions]

Comment: The data collection had the purpose to inform the development of the system, be it through the better understanding of clinical processes, symptoms and issues to track in rheumatic conditions, or patient and clinician perspectives on monitoring.

1.1.2 What types of data will the project generate/collect?

[observational (e.g., sensor data, data from surveys)]

Interview transcripts, survey responses, screenshots, or pictures

1.1.3 What formats of data will the project generate/collect?

[Text files, Numerical, Multimedia]

(1) Interview transcripts: each file will amount to 100KB, up to a total of 10MB; (2) Survey responses: each file will amount to 100KB, up to a total of 10MB; (3) Screenshots or pictures: each file will amount to 5MB, up to a total of 200MB; (4) Reports will amount to 100KB, up to a total of 10MB.

1.1.4 What is the origin of the data?

[Primary data]

1.1.5 What is the expected size of the data?

MB (megabyte)

Comment: (1) Interview transcripts: each file will amount to 100KB, up to a total of 10MB; (2) Survey responses: each file will amount to 100KB, up to a total of 10MB; (3) Screenshots or pictures: each file will amount to 5MB, up to a total of 200MB; (4) Reports will amount to 100KB, up to a total of 10MB.

1.1.6 To whom might it be useful ('data utility')?

[Researchers]

Researchers or industry innovators working on similar contexts to COTIDIANA.

2.1 Reused Data

2.1.1 Will you re-use any existing data and how?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 Will you use metadata to describe the data?

No

3.1.1.3 Will your metadata use standardised vocabularies?

No

3.1.1.7 Will you use naming conventions for your data?

Yes

3.1.1.8 Please provide more details and examples on used naming conversions

Patient interviews: PP[Number of patient participant] Transcript.docx. Clinician interviews: P[Number of clinician participant] Transcript.docx.

3.1.1.9 Will you provide clear version numbers for your data?

Yes

3.1.1.10 Will you provide persistent identifiers for your data?

Yes

Comment: For research publications originated based on the collected data.

3.1.1.11 Persistent identifiers

DOI

3.1.1.12 Will you provide searchable metadata for your data?

No

3.1.1.15 Will you use standardised formats for your data?

No

3.1.1.18 Are the file formats you will use open?

Yes

3.1.1.20 Do supported open-source tools exist for accessing the data?

Yes

3.1.1.22 Will you provide metadata describing the quality of the data?

No

3.1.2 Making data openly accessible

3.1.2.1 Are there ethical or legal issues that can impact sharing the data?

Yes

Comment: The dataset consists of pseudo-anonymised data from a group of patients, clinicians, Pharma. Ethics review board has been sought and Informed Consent obtained from all participants. We are not considering sharing the data externally, apart from selected and processed data in research publications.

3.1.2.2 Will your data be openly accessible?

none

3.1.2.4 How will the data be made available?

[Project website, Repository of Archive]

3.1.2.6 Is the storage sufficiently secure for the data and does the storage provide backup and recovery procedures?

secure with backup and recovery

3.1.2.7 Are there any methods or tools required to access the data?

No

3.1.2.10 Will you also make auxiliary data that may be of interest to researchers available?

no auxiliary data

3.1.3 Making data interoperable

3.1.3.1 Will you use a controlled vocabulary for your data?

No

3.1.3.2 Will you provide a mapping to more commonly used ontologies?

No

3.1.4 Increase data reuse

3.1.4.1 When do you plan to make your data available for reuse?

never

3.1.4.3 Please describe the reason the data will not be made available

Confidentiality of participants and data use purpose defined.

3.1.4.5 Do you have documented procedures for quality assurance of your data?

No

3.1.4.8 Will you provide any support for data reuse?

No

4.1 Allocation of resources

4.1.2 Will you identify a data manager to manage your data, if not who will be responsible for the management of your data?

Yes

4.1.4 How do you intend to ensure data reuse after your project finishes?

[Institutional archive]

5.1 Data Security

5.1.1 What do you plan to do with research data of limited use?

Kept on secure, managed storage for limited time

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on data sharing?

Yes

6.1.2 What are the methods used for processing sensitive/personal data?

[Anonymising data where necessary]

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Considering the DMP is a live document, we have provided the information above in an Excel format within the Consortium shared space so that each member has an overview of the Q&A DMP.