D1.4 Mobilise-D Data Management Plan V2

Mobilise-D

Connecting digital mobility assessment to clinical outcomes for regulatory and clinical endorsement.

Grant Agreement No. 820820

WP1.3 – Data Management Plan

Lead contributor	Mike Jackson (1 - UNEW) mikej@ixscient.com
Other contributors	Brian Caulfield (9 - UCD) b.caulfield@ucd.ie
	David Singleton (9 - UCD) david.singleton@ucd.ie

Due date	30 Sep 2021
Delivery date	30 Sep 2021
Deliverable type	R (Report)
Dissemination level	PU (Public)

Description of Work	Version	Date
	V1.0	30 Sep 2021

Document History

Version	Date	Description
V0.1	28 Sept 2021	First Draft
V0.2	30 Sep 2021	Comments and feedback from workshop
V1.0	30 Sep 2021	Final Version

Table of Contents

1 Executive Summary

This deliverable provides the second version of the Mobilise-D Data Management Plan (DMP) with minor updates to process flows and roles. The project has moved into the second stage (as described below) and the studies and data described within this document are still under development and are subject to change. The deliverable outlines the current understanding of how the research data collected or generated will be handled during and after the Mobilise-D project. It further describes which standards and methodologies for data collection and generation will be followed, and how relevant data will be shared.

2 Introduction and aim

The main objectives of Mobilise-D are threefold: to deliver a valid solution (consisting of sensor, algorithms, data analytics, outcomes) for real-world digital mobility assessment; to validate digital outcomes in predicting clinical outcome in chronic obstructive pulmonary disease, Parkinson's disease, multiple sclerosis, proximal femoral fracture recovery and congestive heart failure; and, to obtain key regulatory and health stakeholder approval for digital mobility assessment.

The Mobilise-D project plan includes a DMP and together with the Consortium Agreement provides the general framework regarding data management, data protection, data ownership, accessibility, and sustainability requirements. As the DMP is an evolving document, some of these aspects may be further described and/or updated in later versions of the document.

In summary, the Mobilise-D DMP gives guidance and provides an oversight of general data management, while each evaluation site needs to provide specific data management information for this plan including, but not limited to, data capture systems, data analysis systems, data protection and data privacy measures, including description of de-identification of data sets and access rules. Project partners do not have to ensure access to parts of research data if such access would compromise their legitimate interests. In such cases, the DMP will contain reasons for not providing access.

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No. 820820. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme, and the European Federation of Pharmaceutical Industries and Associations (EFPIA).

3 General principles

Mobilise-D is a multi-national, multi-disciplinary research programme that involves over 50 collaborators, divided into seven work-packages:

- WP1 Project Coordination & Oversight
- WP2 Algorithm development and technical validation
- WP3 Database development and data management
- WP4 Definition and validation of digital mobility outcomes against clinical endpoints
- WP5 Regulatory, HTA and payer consensus over operational definitions
- WP6 Statistical analysis, evaluation of results and data availability

- WP7 Stakeholder information and results dissemination and exploitation

- All work-packages are tightly integrated so that data, samples and technical expertise are shared throughout the programme. All data from all sites has to be captured according to local institutional guidelines

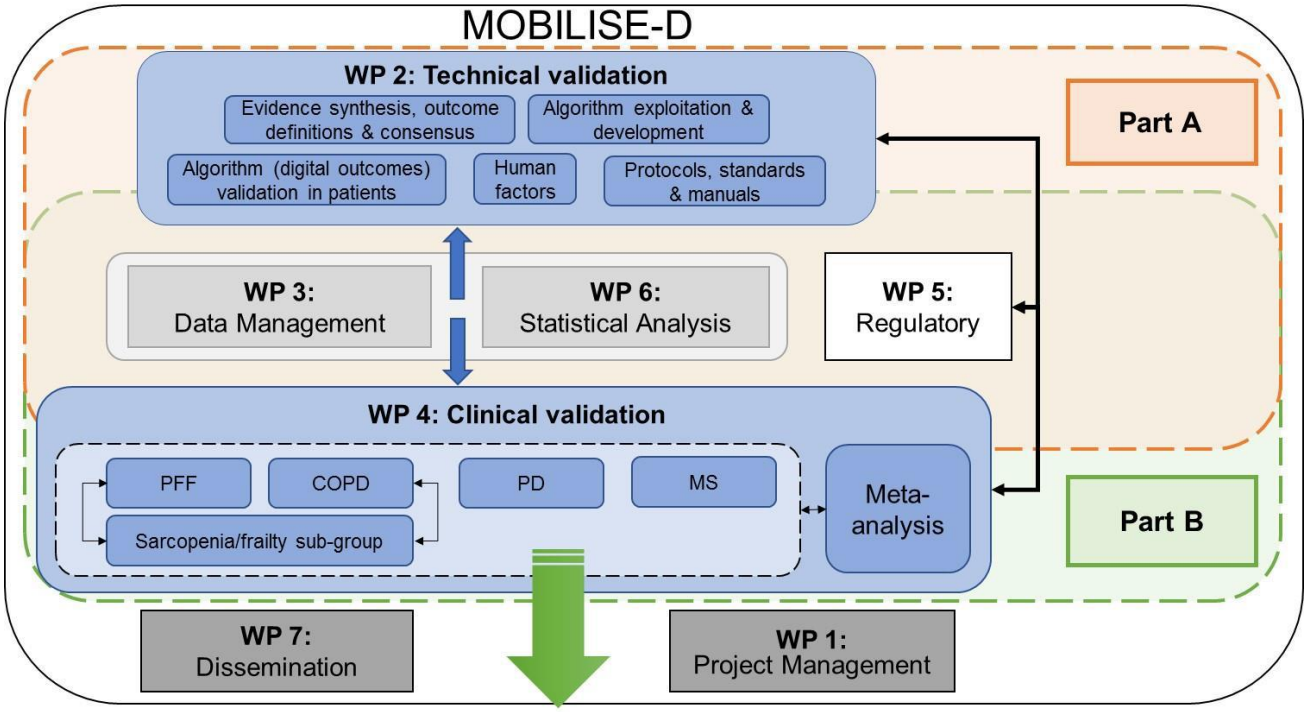

Figure 1: Overview of Mobilise-D Project Structure

Part A - Year 1-2 - Technical validation and qualification advice
Part B - Year 3-5 - Clinical validation and regulatory endorsement

Mobilise-D will generate the following types of data:

- Participant information: anonymised participant demographic and clinical data (such as diagnosis code and disease severity classification)
- Research data: large digital datasets, results from longitudinal cohort studies, results from technical validation studies. This will include wearable motion sensor data, data from ground truth motion capture systems, confounding factor data (e.g. GPS location data), text from semi structured participant interviews, and data from clinical rating and functional assessment scales.
- Evaluation data: results of technical and clinical validation studies; standards (text and spreadsheets)

The DMP follows the principles that research data are findable, accessible, interoperable and reusable (FAIR)[1], as well as being attributable, legible, contemporaneous, original and

accurate (ALCOA)[2]. The general principles on access rules are defined in the Consortium Agreement (Section 8 Intellectual property – Access rights).

4 Compliance with Data Protection Law

By signing the Mobilise-D Consortium Agreement, all member organisations have agreed to comply with all laws, rules, regulations and guidelines applicable to the collection, use, handling, disposal and further Processing of the Personal Data and the Human Samples, in accordance with *Appendix 2: Actions involving Personal Data and/or Human Samples, Consortium Agreement for Mobilise-D*.

Mobilise-D researchers commit to the highest standards of data security and protection in order to preserve the personal rights and interests of study participants. They will adhere to the provisions set out in the:

- General Data Protection Regulation (GDPR)[3]
- Directive 2006/24/EC of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communication services or of public communications networks[4]
- Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)[5]
- Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data[6]

Prior to collecting, storing, and processing sensitive personal data, the consortium will seek approval of the applicable local data protection authorities. Consent forms will contain information on how personal data will be managed. To secure the confidentiality, accuracy, and security of data and data management, the following measures will be taken:

- All personal data obtained in Mobilise-D will be available to partners within the consortium only after anonymization. Keys to re-identification will be held confidentially within the respective research units. In situations where re-identification of study participants becomes necessary, for example the collection of additional data, this will be possible only through the research unit and in cases where informed consent for such cases has been given.
- Personal data are entered to secure websites. Data are processed only for the purposes outlined in the patient information and informed consent forms. Use for other purposes will require explicit patient approval. Also, data are not transferred to any places outside the consortium without patient consent.
- None of the personal data will be used for commercial purposes, but the knowledge derived from the research using the personal data may be brought forward to such use as appropriate, and this process will be regulated by the Grant Agreement and the Consortium Agreement, in accordance with any generally valid legislation and regulations.

5 Overview of data managers, data repositories and access rules

The responsible roles for the repositories used in the Mobilise-D project are listed in Table 1 below. The platform “e-Science Central” (e-SC) is used as the data management platform for Mobilise-D members and will be used to upload and receive pseudo-anonymised datasets, reports and algorithms. The final version of the DMP will have details regarding the long-term storage and preservation of anonymized data particularly after the end of the Mobilise-D project. The relevant committee will determine where it is allowed to store data and under which license(s). Datasets containing personal data in the possession of partners other than the research data owner must be destroyed at the end of the Mobilise-D project. Other non-public and public datasets not containing personal data will be stored for the recommended timeframe from the end of the Mobilise-D project to ensure their long-term availability to future researchers.

Role	Responsibilities	e-Science Central	DW
		Access Levels	
Study Volunteer	N/A	Request personal data as per GDPR	N/A
Clinical Investigator	Acquire and upload data	RW	-
Data Controller (Lead CI)	Data protection and processing at sites	RW	R
Data Manager	Authorises access and governs data management policies	RW	R
Platform Manager	Provide administration of and access to platform	All	All
Data Analyst	Analyse/interpret data and develop data optimization strategies	R	RW
Statistician	Responsible for implementation of pre-defined statistical analyses on the derived data and results files.	-	RW

R=Read , W=Write, All=access to all data

Table1: e-SC access roles

All questions related to data management such as rules for uploading data, requests for access rights should be sent to the data manager.

All staff who are tasked with collecting data are required to complete requisite training. Scheduled training webinars are provided by content experts and a training log is maintained by the project administrator. All training material is also made centrally available on SharePoint along with recordings of training webinars. When training is completed the site administrator can then submit a site staff checklist signed by their PI requesting access to the relevant platforms. The data manager will approve/reject the requests as required and contact the platform managers to create these accounts. Data collection/ingestion on e-SC is organised into e-SC studies based on site and disease cohort. For example, we are recruiting MS patients in Sheffield (USFD) and we have created a study MS-USFD within which we create patient IDs (pseudonyms) and store all of the relevant patient data. When access requests are approved, site staff can be added to that study based on their role.

The platform manager creates studies for all site/disease-cohort combinations (see Appendix B) and allocates a patient ID range to each. Therefore, a patient ID is auto-generated when the assessor creates a new patient ID. When a patient is created on e-SC, the ID is entered into the e-SC database and a patient folder is created on the e-SC file server. All data manually uploaded is stored in this folder along with patient data transferred via API (e.g. sensor data). The platform manager provides access tokens to relevant 3rd parties to use the e-SC APIs. All data entered through web forms is stored as events in the e-SC db associated with that same patient ID. If any erroneous data needs to be corrected a Data Change Request (DCR) form must be completed by the site and approved by the data manager before being implemented by the platform manager. Only the platform manager can modify the e-SC database entries.

e-SC serves as the main ingestion hub for all data, including electronic Patient Reported Outcomes (ePRO), Clinician Reported Outcomes (eCRO), Case Report Forms (eCRF), sensor data, and confounding factor data. However, this data is in very different formats, some in a database, other on file servers and not easy to interrogate. The platform manager in collaboration with the statistical analysis team has developed a data warehouse, and Extract, Transfer and Load (ETL) processes to centralise all this data in one place and in one format. The ETL processes are a series of scripts used to extract the data from e-SC, parse the data and load into the data warehouse (normalized database structure). Final analysis of the datasets will be completed via access to the data warehouse and this will be provided by the platform manager based on the necessary roles.

6 Overview of data generated and collected in Mobilise-D

Mobilise-D members use a SharePoint platform to facilitate collaboration between members, to plan deliverables, to track progress of all tasks, and to store meeting minutes and task reports.

The research data generated and collected during the Mobilise-D project can be divided into two categories;

1. Technical Validation Study (WP2 / WP3)
2. Clinical Study (WP4 / WP6)

Patient data will be generated and processed during the activities planned in WP2, WP3 and WP4. All data will be uploaded to the Mobilise-D platform e-SC hosted on AWS. During the technical validation study some data (CRF/sensor data) may be uploaded manually via a portal provided. All other data will be transferred via API from 3rd party servers. No Personal Identifiable Information (PII) will be uploaded to e-SC. See Figure 2 below;

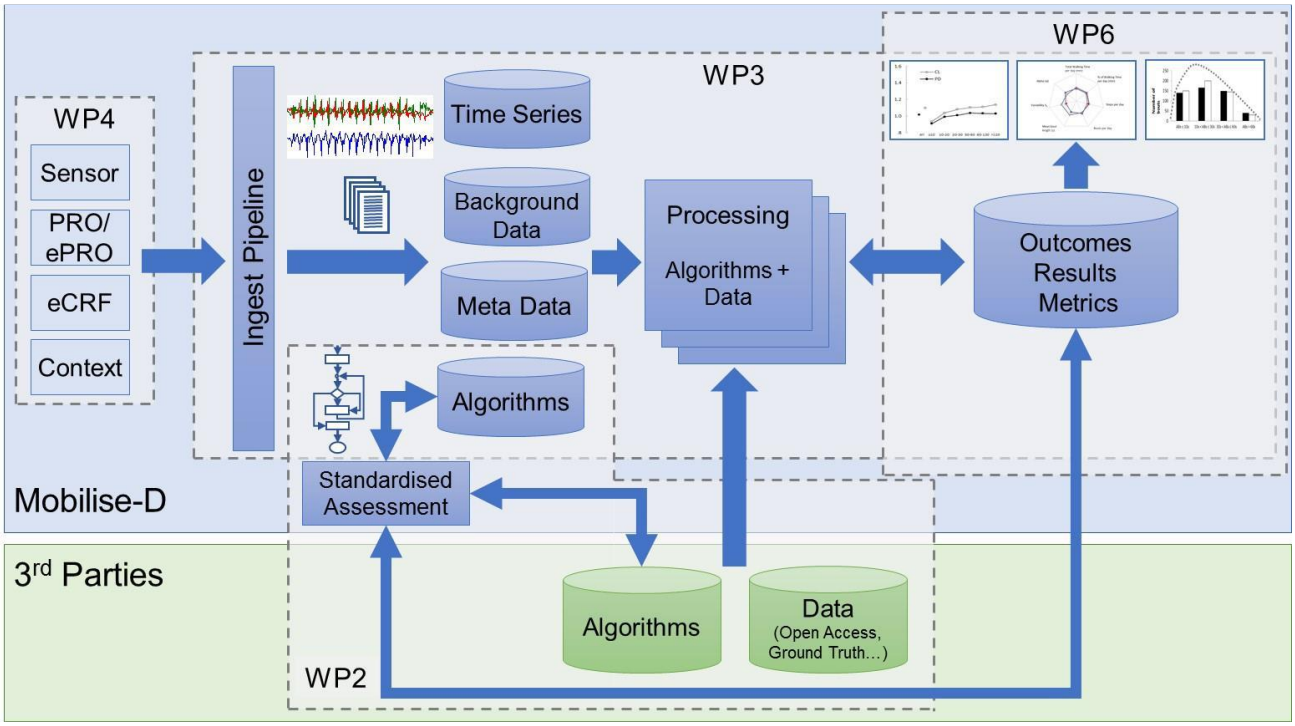

Figure 2: Overview of Mobilise-D Data Management

6.1 Technical Validation Study (WP2/3)

6.1.1 Description of the data

For the technical validation study existing digital mobility data from partners in the consortium will be synthesised into a metadatabase of real-world and laboratory data and algorithms to inform validation. Using the Dynaport Movemonitor+ from technology partner McRoberts, WP2 will produce a validated device-algorithm pair, and associated technical, clinical and patient specific standards for clinical validation. 5 sites will participate in the recruitment and assessment of participants in the technical study. Data will be collected from study volunteers over 9 days using a combination of clinical, laboratory, and home assessments. These activities are illustrated in the flowchart below;

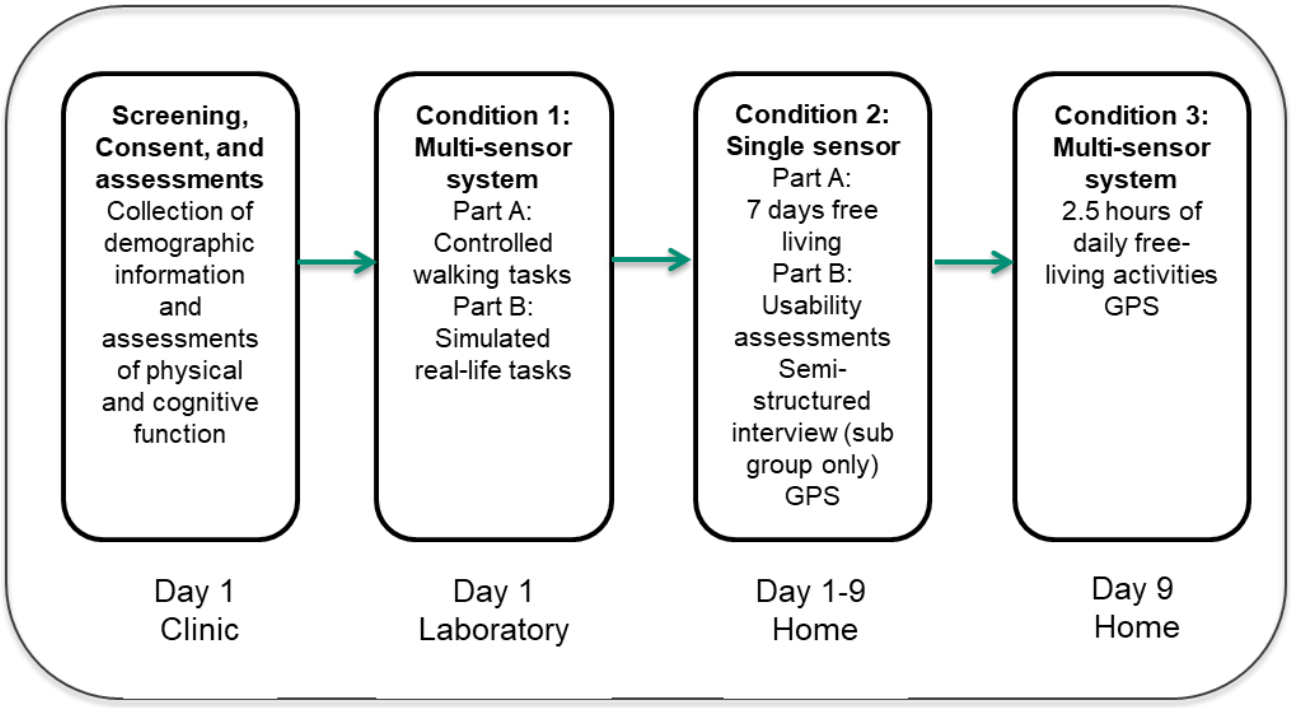

Figure 3. Technical study activities, time points and locations.

Data set reference and name

The broad umbrella term used to describe the data collection comprising multiple datasets in the technical validation study is **Mobilise-D-TVS**. This collection will include the following datasets;

- Demographic, cognitive, physical function
- Multi-sensor-lab data
- Single-sensor data
- Multi-sensor-home data
- Patient Reported Outcomes
- Semi-structured interviews
- GPS annotated data
- Legacy data

6.1.2 Data collection/generation

Day 1: Clinic

Assessment data is collected by clinical practitioners using paper-based and/or eCRFs and collated in e-SC.

Day 1: Lab

Data is collected from a number of sensors as follows;

- **Optolectronic stereo-photogrammetric** systems use cameras to capture the human body motion by tracking the trajectories of spherical retroreflective markers attached

to the body. Only the marker-trajectory information is relevant for Mobilise-D. This information will be exported as a .csv file and manually uploaded to the e-SC platform. No video files will be uploaded to e-SC however annotations of those files may be uploaded.

- **INDIP system** includes 4 Magneto-Inertial Measurement Units (MIMU), 2 distance sensors and 2 pressure insoles). Each MIMU has an on-board flash memory and can record at least 5 hours of acquisitions at 100Hz. A Windows-based application is used to manage the INDIP system (including saving data at the end of the assessment). The number of files is dependent on the protocol. If 5 tests then 5 x 4 files (1 for each MIMU for each test). The format of each file is .txt. Each file is manually uploaded to e-SC via the portal provided.

- **DynaPort MM+** (McRoberts MoveMonitor) consists of a small and light casing containing a tri-axial accelerometer, tri-axial gyroscope, rechargeable battery, USB connection and 1GB flash memory. The sensor will be worn at all times (except bathing) and will store 7 days of data. The sensor is then physically connected to a PC and McRoberts software is used to transfer this data to the e-SC platform. Both raw and processed data can be transferred to Mobilise-D.

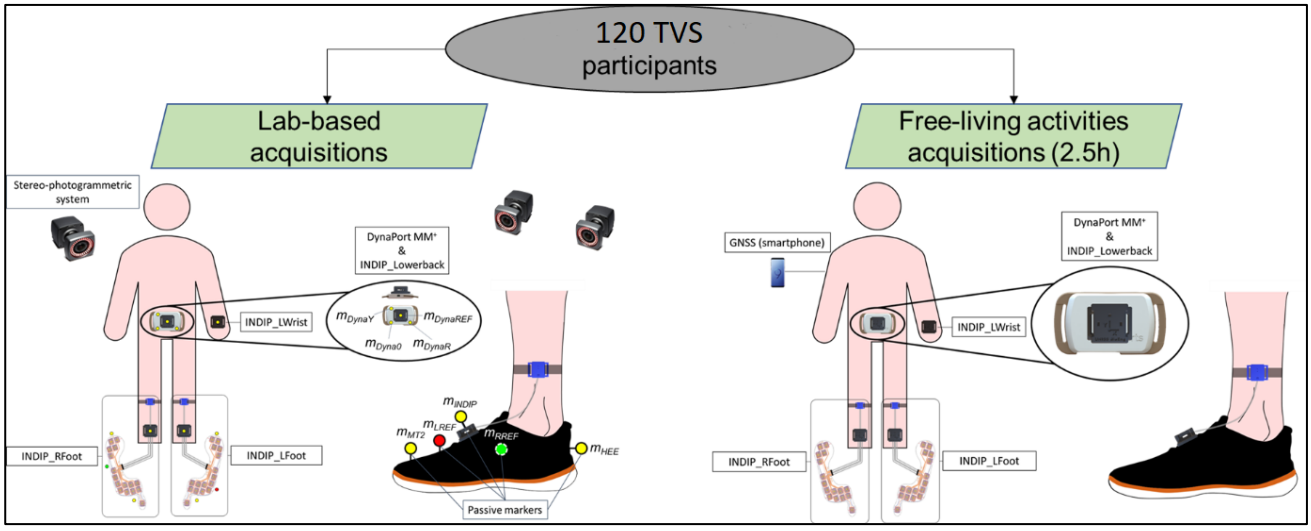

Figure 4. TVS setup for in-lab and out-of-lab assessments

Day 1-9: Home

This period will include 7 days of free-living activities wearing the Dynaport MM+ sensor, along with a mobile phone (BYOD or provisioned) which will be used to collect patient reported outcomes (ePRO), and GPS data to understand confounding factors. Researchers will visit the study volunteer and reclaim sensor data which will be uploaded as described previously, ePRO data will be uploaded to the ePRO partner (ERT) platform and from there de-identified data will be sent to e-SC via API. GPS coordinates will not be sent to e-SC but de-identified processed data will be sent from GPS server to e-SC via API. This data will be in JSON format. Human factors will be addressed using a combination of quantitative (e.g. deployment of instruments) and qualitative (e.g. semi-structured interviews) techniques in a sub-sample of the users involved in the seven days validation protocol. The audio files from these interviews will be transcribed and transcriptions will be manually uploaded to e-SC.

The audio files will be securely retained at each site until project completion after which point they will be destroyed (see Appendix A).

Day 9: Home

This is another series of activities in a free-living environment conducted over a 2.5 hour period. During this time the study volunteer will wear the Dynaport MM+ sensor, the INDIP system, and carry the phone for GPS related information. All data will be transferred as described earlier.

Data Tools used

Mobilise-D will produce different types of software. Algorithms will be developed/refined to process the raw signals from the sensors and extract from them specific quantification of mobility such as average step length, average step cadence, or average walking speed. Source code will use standard programming language format files, depending on the chosen language for the development. All code will be developed using SDLC and hosted on a shared SCM platform. Tools used for data visualization will be determined at a later stage.

6.1.3 Data organisation, documentation and metadata

For version one of the Mobilise-D data processing platform, we will develop a simple data model for the project data sets and deploy a basic website that authorized project members can use to upload, browse and process data. In its initial phases, the Mobilise-D project needs to gather together existing sets of Gold Standard data and make use of a set of published algorithms to analyse them. Data will be organised hierarchically within the data model in order to group together sets of data representing different collections, measurement types etc.

Data for evaluation.

Legacy datasets and algorithms will be provided by Mobilise-D partners. Existing public datasets (e.g. weather) may also be used.

Research data and metadata

Quantitative data will be generated during the technical validation study in clinical, laboratory, and home environments utilizing multiple sensors and a ground truth reference system. Qualitative data will be collected in the form of questionnaires and interviews with end-users who will test the Mobilise-D tools. Data dictionaries will be provided at a later stage describing the variable names/labels etc.

Processed data files

Source code and research data will be accompanied by a readme file including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed. Documentation will also include details on the methodology used, analytical and procedural information, any assumptions made, and the format and file type of the data. In the case of software it may also include installation instructions and usage examples.

Data Format

Semi-structured interview data and system annotations will be collected in text files in DOC, PDF, or TXT format. Spreadsheet data will be collected in CSV format. Key/Value pair data will be collected in JSON format.

File naming convention

Files will be named according to their content to ease their identification. Software development will be managed through specific software versioning and revision control systems (SCM) such as git.

Each file will be labelled in standardised format, including information on:

- Centre
- Participant unique ID
- Data source/ modality
- Time point – YYYYmmdd

Example
centre#-patient#-testname-ddmmyyyy.extension

E.g.
10-20000-moca-22092020

6.1.4 Data storage and security

Managing, storing and curating data

During the project both software (algorithms/applications) and data will be stored on the data management platform e-SC built on AWS. Source code however will be managed through Gitlab VCS. Data will be encrypted in transit (HTTPS) and at rest. Flat files will be stored in S3 buckets which are encrypted using AES-256 encryption. RDS databases are also encrypted using AES-256. Platform managers will be responsible for the data while in the data management platform. Amazon RDS creates and saves automated backups of the DB instance securely in Amazon S3 for a specified retention period. In addition, we will create daily snapshots which are kept until we explicitly delete them.

The data stored in the Mobilise-D platform is pseudo-anonymised by using a unique identifier for every participant. Technical implementation to combat specific data security issues used within the deployment of the web service are:

1. e-SC runs on virtual servers in AWS. The virtual server platform provides ability to add resources automatically depending on demand (auto-scaling).

3. The webservers are configured using security. All servers are secured using a firewall technology which has automated blocking of non-legitimate access attempts.

4. General shell/admin-level access to webservers is restricted to a primary and backup administrator - there is no general access to project members.

5. Access to the data stored within the PostgreSQL is restricted to the primary and backup administrators only.

6. Access to the website is restricted to a whitelist of staff managed by the platform managers.

7. Access/View/Editing of data records is performed over SSL to ensure the integrity of data from client browser to database record store and is restricted to a number of staff in each partner in the project. Only those authorized as per Table 1 can view records across multiple sites.

6.2 Clinical Study (WP4/6)

6.2.1 Description of the data

The MOBILISE-D clinical study is a longitudinal (non-interventional) observational cohort study conducted in ten different European countries in 16 different partner sites. Participants will have a baseline assessment and follow-ups (every 6 months over a period of 24 months – see Table 2) for a duration of 7 days each using a wearable inertial sensor. The device will be placed at each visit and returned for subsequent data download to a local clinic PC. This data is then automatically uploaded to McRoberts platform and raw data will be sent via an API to the e-SC platform developed in WP3. Participants will also complete ePROs and this data will be sent automatically to the ERT servers and from there will be pushed to e-SC. Environmental data will also be captured and uploaded to e-SC to identify confounding factors.

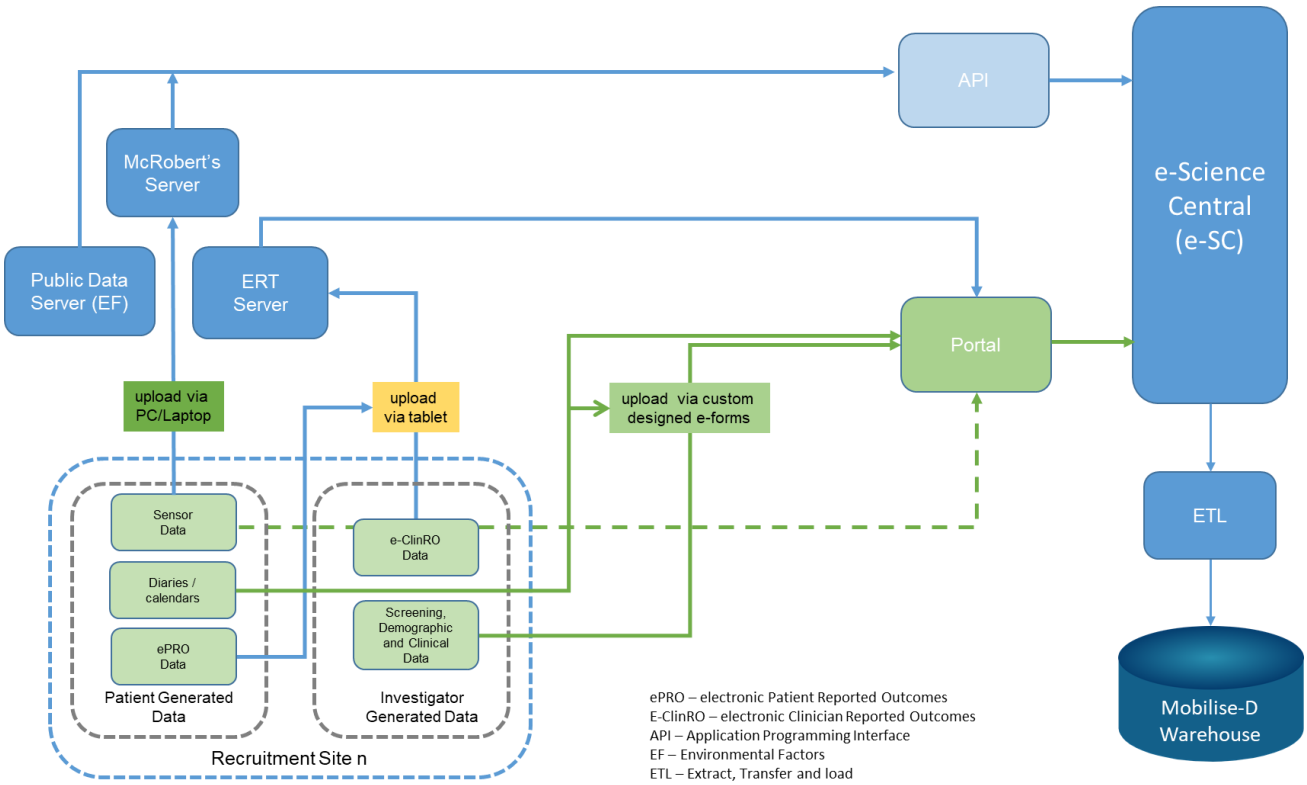

Figure 5. Clinical study data path(s).

Data set reference and name

The broad umbrella term used to describe the data collection comprising multiple datasets in the clinical validation study is **Mobilise-D-CVS**. This collection will include the following datasets;

- Demographic, cognitive, physical function
- Single-sensor data
- Patient Reported Outcomes
- Semi-structured interviews
- Environmental data

M1	M6	M12	M18	M24
BL (T1)	FU (T2)	FU (T3)	FU (T4)	FU (T5)
LV	LV	LV	LV	LV

Table 2. Schedule of assessments per patient. BL=baseline; FU=follow-up.

6.2.2 Data collection/generation

The baseline visit (T1) will be performed in the clinic. All follow-up visits (T2-T5) may be offered as a home visit if the participant is not able or willing to attend a hospital visit. The reason for conducting the home visit must be documented. Subsequent visits are to be completed in clinic again, unless participants are still unable or unwilling to attend a hospital visit. Data will be collected using eCRF at baseline and uploaded to e-SC for data aggregation. Ongoing patient reported outcomes will be collected via ePRO using the ERT platform and this data will be uploaded to e-SC. Raw data from McRoberts sensor for the 7-day wear periods will be pushed to e-SC, along with contextual data from public data services.

6.2.3 Data organisation, documentation and metadata

All data transferred to the e-SC platform from the automated pathways will be stored in S3 on AWS. Pointers to this data will be maintained in a PostgreSQL database. Analysis tools and data storage will be configured to provide access to the derived data. This will be further outlined in the final DMP.

Data for evaluation.

N/A.

Research data and metadata

Quantitative data will be generated during the clinical validation study in clinical, and home environments utilizing a single sensor and ePROs. Qualitative data will be collected in the form of questionnaires and interviews with end-users. Data dictionaries will be provided at a later stage describing the variable names/labels etc.

Processed data files

Source code and research data will be accompanied by a readme file including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed. Documentation will also include details on the methodology used, analytical and procedural information, any assumptions made, and the format and file type of the data. In the case of software, it may also include installation instructions and usage examples.

Data Format

Semi-structured interview data and system annotations will be collected in text files in DOC, PDF, or TXT format. Spreadsheet data will be collected in CSV format. Key/Value pair data will be collected in JSON format.

File naming convention

For all manual data collection in the clinical study (e.g. falls diaries) the file naming convention will follow that outlined in the technical validation study. All other data collection is automated and uses the participant ID as the unique identifier.

6.2.4 Data storage and security

Managing, storing and curating data

During the project both software (algorithms/applications) and data will be stored on the data management platform e-SC. Source code however will be managed through a Gitlab VCS. Data will be encrypted in transit (HTTPS) and at rest. Flat files will be stored in S3 buckets which are encrypted using AES-256 encryption. RDS databases are also encrypted using AES-256. Platform managers will be responsible for the data in the data management platform. Amazon RDS creates and saves automated backups of the DB instance securely in Amazon S3 for a specified retention period. In addition, we will create daily snapshots which are kept until we explicitly delete them.

The additional storage facility (Data Warehouse) for derived data will be fully explored in later versions of the DMP. See Appendix C.

7 Data sharing and re-use

The Consortium is aware of the mandate for open access of publications and research data. As previously mentioned, the Consortium will choose the appropriate scientific publication and data repository for the project outcomes. The Consortium will ensure that scientific results that will not be protected and can be useful for the research community will be duly and timely deposited in the scientific results repository. This will include;

- Electronic copies of the final version or final peer-reviewed manuscript accepted for publication.
- Project public presentations and any other kind of dissemination material.
- Research data needed to validate the results presented in the deposited publications

7.1 Procedures for making data findable

The information collected and updated via Appendix A will be available to the Mobilise-D consortium through e-SC during the project and the public through a public repository determined at a later stage of the DMP. The datasets will receive unique identifiers as a result of uploading to the relevant repository. This will enable the easy identification of datasets available and identify the data owner.

7.2 Re-use of Mobilise-D results by third parties

For those external individuals/institutions wanting to use Mobilise-D generated or collected data during the course of Mobilise-D, a data sharing policy has been defined and approved within the project. The section below outlines the key elements of this.

Data Sharing Principles

Mobilise-D recognises and will follow the principle that datasets generated should be made available to parties outside of the consortium, for research purposes, as soon as reasonably practical.

In addition, Mobilise-D recognises the hard work of the consortium in generating the data associated with the project. Further, Mobilise-D recognises its commitment under the Grant Agreement and Consortium Agreement (section 7) to disseminate the results of the project

as soon as possible, including algorithms. However, dissemination can be delayed if it goes against the legitimate interests of a beneficiary (CA section 7.5.1.1).

Mobilise-D recognises the needs of academic partners within the project to publish, in peer reviewed publications, results generated from the project to maximise project impact and their own academic standing and as such the release of data can be legitimately deferred to enable this.

In addition to this, EFPIA partners with Mobilise-D are investing their own time and resources into the project and have concerns that making data and results generally available before they have had time to work with the data, may give competitors, who have not made such an investment, an unfair commercial advantage.

Thus, a balance needs to be struck between the legitimate interests of the consortium and the principle of making data available as soon as reasonably practical.

The Exploitation and Impact Sub-committee will oversee the data sharing process and determine when data can be made available for third party actors.

As a matter of policy data can be released when:

- The data within the dataset has been fully collected,
- Quality checked, analysed and curated.
- Have been used by Mobilise-D in peer reviewed publications
- Has been agreed by the consortium it can be released

Data Sharing Commitment

Mobilise-D commits to:

1. Making publicly available a list of datasets and metadata generated within the project.
2. Sharing summary study results with patients who have participated in our studies.
3. Make generally available the validated algorithm(s) used to develop DMO's following academic publication.
4. Sharing with parties external to the project datasets and other data in a reasonable timeframe via an authorisation procedure.

Ownership of Data

The Mobilise-D Consortium Agreement stipulates that the participants who generate data own that data. For most data generated within Mobilise-D, multiple project participants will be involved so multiple participants will all have co-ownership of the data. Mobilise-D recognises the requirement to make data available externally and that ownership of data should not be an impediment to this.

Authorisation for Data Release

The Mobilise-D Steering Committee will be responsible for overseeing data sharing via the Exploitation and Impact Sub-committee (EIS). It will, on a regular basis, review project data and its availability. Where appropriate and subject to the data sharing principals defined in section above, it will recommend the release of data, subject to authorisation via the Data Sharing Authorisation Procedure and the Data Sharing Methodology defined below.

The Data Sharing Authorisation Procedure enables the data proposed for release to parties outside the consortium to be presented to all beneficiaries and for them to object if their legitimate interests are damaged because of this. This process is required to be compliant with the terms of the Consortium Agreement.

A situation may arise where an external actor may request access to data before it has been authorised for release. Such a request will be reviewed by the EIS. If the purpose for data access overlaps with the scope of Mobilise-D or the effort to extract the required data is significant then the EIS can reject the application. Otherwise, the EIS can seek approval from beneficiaries to release data or partial data on a case-by-case basis and enact the Data Sharing Authorisation Procedure.

Data Sharing Authorisation Procedure

When the EIS identifies that a data is available for release to third parties the following procedure will be enacted.

The EIS will contact all participants within Mobilise-D to inform them of the proposal to release a dataset. This will include the name and description of the dataset and where feasible metadata and data sample.

Participant organisations have 30 calendar days to object to the release of data giving reasons and justification why their legitimate interests would be harmed. Included in this response must a proposed revised date by which data can be released.

- The EIS will review this request.
- If the EIS agrees with the delay a new release data will be set.
- If it disagrees with the objection it will work with the objecting participant to resolve the matter. If no resolution is found the matter will go to the Mobilise-D Steering Committee for resolution.

Data Sharing Methodology

Mobilise-D will keep a log of data generated within the project. This will include the name of the data set, when it was completed, project publications / results arising from the data, what work within the project is still using that data and an estimate of when it will be more widely released.

A short form of this log will be made available on the Mobilise-D website.

Where possible data samples will be made available via the project website to highlight to potential users what is available in the data set.

An application process and associated forms will be available on the website. This will coincide with the first approved data release.

Mobilise-D recognises that there are costs, not foreseen in its Grant Agreement, associated with making data available to actors outside of the project. These include:

- Storage and maintenance
- Processing of data access applications
- Generation of data subsets from access requests

As such, Mobilise-D reserves the right to charge a fee for access to the data.

A range of data sharing mechanisms have been investigated, ranging from free access via a public repository to access via an Analytical Environment. Each mechanism has associated pros and cons as well as costs.

Mobilise-D will make data available via a data sharing agreement to external actors via an application process. Once this process has been completed and any relevant fees are paid, the specific data requested will be accessible from the eScience central platform used by Mobilise-D to collect and curate data.

This methodology may be changed by agreement within the consortium. In addition, provision for access after the end of the project needs to be considered. This is discussed in the Sustainability section below.

The Data Sharing Agreement will be generated by Newcastle University. Key elements of this are highlighted terms and conditions section below. Mobilise-D will define the agreement depending on the data requests, proposed use and requesting party.

Data Access Application Process

The following data access process will be applied:

Step 1: If necessary initial informal discussions will be held between the applicant and Mobilise-D to discuss the access requirements. A Non-Disclosure Agreement may be put in place to enable this.

Step 2: An application form for access to Mobilise-D data must be completed. This may include among other items, the institution, researchers, collaborators, contact details, data requested, funding source and purpose / use of the data.

Step 3: Mobilise-D will review the application and contact the applicant with any feedback or request for additional information. In addition, a Data Sharing Agreement will be sent to the applicant's institution for agreement. The terms of this agreement may be dependent on factors such as institution type and the purpose of the data usage etc.

Step 4: On agreement of the terms of the Data Sharing Agreement and approval of the application, an access fee request will be raised, if applicable.

Step 5: Following payment of the access fee and return of the signed Data Sharing Agreement, Mobilise-D will inform the applicant by when the requested data will be released and how to access it.

Data Sharing Application

Applicants must submit a Data Sharing Application using the form available on the Mobilise-D website. An appointed person from Mobilise-D will act as the contact person for that applicant.

All applicants must specify the intended funding source for the work using Mobilise-D data. In addition, applicants must state whether the proposed work is partially or completely based on the use of Mobilise-D data. Furthermore, all applicants should mention any financial conflict of interest or any relationship with commercial partners.

If applicants for data use wish to apply for funding once the application has been approved, and the grant application references Mobilise-D data, Mobilise-D must be informed about the funding application.

Applications for use of data within a student project must also use the application process.

Evaluation of the Application

Mobilise-D will evaluate the application and decide whether to give access to the data, in which form, and for how long, and inform the applicant via email. The decision to provide access may also include specific terms to be included in the data sharing agreement.

Approval is subject to the applicant's signing and returning the data sharing agreement and this being counter signed by a duly appointed Mobilise-D representative.

If the proposed use of the data is subject to a grant application, the data release, if approved, will be subject to verification that the grant application has been successful.

Use of the data outside those uses specifically stated in the application form will require the applicant to submit a new application and then Mobilise-D to issue a new approval.

If the application is rejected, the applicant will receive an explanation of the rejection. Applicants can appeal the decision within one calendar month of rejection notification. If the appeal is rejected, then the applicant must re-apply for access.

If the application is successful, Mobilise-D will endeavour to provide the dataset(s) and associated documentation within 2 calendar months.

Terms and Conditions

Key terms and conditions are provided below, however additional terms may be included within any data sharing agreement.

Data application

Only the main investigator of the project is entitled to apply for data access.

- The applicant must conduct high quality, ethical research when using Mobilise-D data.
- The description of the project, analysis and publication of data must be as specific and focused as possible.
- Major changes to the project will require a new data sharing application.

- If the proposed use of the data is subject to a grant application, the data release, if approved, will be subject to verification that the grant application has been successful.

Data Sharing Agreement

Once the Data Sharing Application has been approved, Mobilise-D will provide the applicant with the terms and conditions of the data sharing agreement. This must be signed by a duly appointed representative of the applicant's institution and counter signed by Mobilise-D, together with receipt of any applicable fees, before the data can be released and used.

Ethics and confidentiality

Applicants must provide a copy of the ethics approval of their research project before data is provided. Ethical approval from the data user's local ethical committee is the responsibility of the data user. If necessary, for work defined within the proposed application, data can be provided without restriction within the European Economic Area (EEA). Mobilise-D data can also be transferred to a country or territory outside the EEA if the applicant and their collaborators provide an adequate level of protection of personal data and operate under the data protection scheme in place in their country. If requested by Mobilise-D, a copy of this data protection registration should be provided.

All Mobilise-D data are anonymised or pseudonymised before sharing. It is forbidden to match or attempt to match individual records to any other data. Additionally, it is restricted to copy the data or send it to anyone else.

The main applicant and their collaborators must abide by these terms and conditions. Failure to comply with them will result in Mobilise-D's refusal of any future data sharing applications from the main applicant and/or their collaborators.

Data usage and security

- Data users must ensure that no participant's identity is disclosed under any circumstances.
- Data users must NOT try to de-anonymize the data by any technique or to match or attempt to match individual records to any other data.
- The data users must not copy the data or send it to anyone else outside the terms of the data sharing agreement.
- Data must be used only for the specified project and within the timeframe specified in the application form. Major changes require a new application.
- Once the project has been approved, data users should work under the data protection scheme that operates in their country. If requested, a copy of this data protection registration should be sent to Mobilise-D.
- Mobilise-D must be informed of any new collaborators in the project.
- Data cannot be transferred to persons not working on the project described in the application. Data can only be transferred to persons listed in the application form.

- Data users will not have sole and exclusive access to their required set of data.
- Analyses are allowed only according to the protocol described in the application.
- Data errors must be notified to Mobilise-D.
- Secure data access, such as passwords, firewalls, etc., must be in place to ensure that the data are kept secure.
- Data may not be stored on servers or cloud storage where terms and conditions of use may enable a 3rd party access or ownership of data.
- After the project duration mentioned in the proposal, data should not be used anymore and must be deleted. The applicant must complete a document where they state the data have been deleted.

Publications

- The scope of research projects should ideally be aimed at the publication of scientific articles within two years after receipt of the requested dataset.
- The name of the Mobilise-D study must be included in the title or subtitle.
- If the project has received significant input from Mobilise-D, then the person providing that input from Mobilise-D may also be included as an author.
- Mobilise-D should receive credit for the vast amount of work carried out to establish the resource data. Therefore, a suitable note of acknowledgement should be agreed with Mobilise-D and included in publications. This acknowledgement must also include that Mobilise-D was funded under the Innovative Medicines Initiative. This statement will be provided in the data sharing agreement.

It is expected that Mobilise-D will be updated on the progress of the project and will be sent all material accepted for publication.

Sustainability

Mobilise-D is cognisant of the fact that it is not in itself a legal entity and that the project has a finite lifetime. The project also recognises that the methodologies described within this document will not be able to be implemented in their current form once the project ends, although there will be ongoing responsibilities on participants from the Consortium Agreement.

Before the end of the project, Mobilise-D will aim to approve all available data for release and authorise the Digital Health Catalyst set up by Mobilise-D to maintain and manage access to the date generated within the Mobilise-D project. This will require the Digital Health Catalyst to become a recognised legal entity.

8 Ethical aspects

Patient Information and informed consent procedures will be approved by the relevant local ethics boards. Data collectors collecting personal data will inform the study participants about the project in an appropriate manner, including:

- the identity of the data controller
- the voluntariness of the collection of data
- the purposes of the processing
- the nature of the processed data, including its type (identifiable, coded, anonymised)
- the handling of the data
- the existence of the right of access to, and the right to rectify the data concerning themselves
- the sharing of data across research groups
- that consent may be withdrawn and how this is done

There are no other ethics issues currently identified beyond those discussed above. Any potential issues that arise during the project duration will be presented to the Ethics and Data Monitoring Committee who will ensure they are addressed by taking the appropriate organisational, legal, and regulatory steps.

9 List of abbreviations

ALCOA	Attributable, legible, contemporaneous, original and accurate
API	Application Programming Interface
AWS	Amazon Web Services
CA	Consortium Agreement
DMP	Data Management Plan
EU	European Union
ePRO	Electronic Patient Reported Outcome
eCRF	Electronic case report form
EFPIA	European Federation of Pharmaceutical Industries and Associations
FAIR	Findable, accessible, interoperable and reusable
e-SC	e-Science Central
GDPR	General Data Protection Regulation
GPS	Global Positioning System
IMI	Innovative Medicines Initiative

SDLC	Software Development Life Cycle
SCM	Software Configuration Management
WP	Work Package

10 Conclusions

A robust Data Management Plan improves the understanding of the data to be generated within a project, and of the requirements of securing and archiving that data. It also highlights the data publication potential of a project – data sets which can be released are identified early and appropriate steps can be taken to ensure that sharing happens in a timely and efficient manner. This document describes the development of just such a DMP for the MOBILISE-D project. This plan is a key element in maximising the impact of MOBILISE-D. The data management plan is not a static document, but will continue to undergo review and formal revisions as MOBILISE-D progresses.

11 References

[1]: The European Commission D-GfRI. H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. 2016. Available from https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[2]: Guidance for Industry, Electronic Source Data in Clinical Investigations. U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH); 2013 [cited 15 February 2017]. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691.pdf

[3]: Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation); 2016 [cited 15 February 2017]. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN.

[4]: Directive 2006/24/EC of the European Parliament and of the Council on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC; 2006 [cited 15 February 2017]. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006L0024&from=en.

[5]: Directive 2002/58/EC of the European Parliament and of the Council concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications); 2002 [cited 15 February 2017].

Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32002L0058&from=en.

[6]: Directive 95/46/EC of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data; 1995 [cited 15 February 2017]. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en.

Appendix A Data Table

Collected or Created	Title	Description	Category	Type	Format	Size	Owner	Privacy	Storage / Storage for public access	Backup frequency	Destroyed at end of project	Duration of preservation (in years)	Work Package
	Mobilise-D-TVS-Clinic-Base												
	Mobilise-D-TVS-Lab-SP												
	Mobilise-D-TVS-Lab-INDIP												
	Mobilise-D-TVS-Lab-DMM												
	Mobilise-D-CVS-Home1-DMM												
	Mobilise-D-CVS-Home1-GPS												
	Mobilise-D-CVS-Home1-ePRO												

Appendix B CVS e-SC Studies

Site #	Site Acronym	Patient ID range	Study Name (e-SC)	Study Code (e-SC)	Cohort
10	CAU	10000-10999	PD-CAU	PD10	PD
11	CHUM	11000-11999	PFF-CHUM	PFF11	PFF
12	ICL	12000-12999	COPD-ICL	COPD12	COPD
14	KUL1	14000-14999	COPD-KUL1	COPD14	COPD
15	KUL2	15000-15999	PD-KUL2	PD15	PD
16	NTNU	16000-16999	PFF-NTNU	PFF16	PFF
17	PFLG	17000-17999	COPD-PFLG	COPD17	COPD
18	RBMF	18000-18999	PFF-RBMF	PFF18	PFF
19	TASMC	19000-19999	PD-TASMC	PD19	PD
20	TFG	20000-20999	COPD-TFG	COPD20	COPD
21	UKER	21000-21999	PD-UKER	PD21	PD
22	UNEW	22000-22999	PD-UNEW	PD22	PD
23	UNN	23000-23999	COPD-UNN	COPD23	COPD
24	USFD	24000-24999	MS-USFD	MS24	MS
25	USR	25000-25999	MS-USR	MS25	MS
26	UZH1	26000-26999	COPD-UZH1	COPD26	COPD
27	UZH2		COPD-UZH2	COPD27	COPD
28	UZH3		COPD-UZH3	COPD28	COPD
30	ISG1	30000-30999	COPD-ISG1	COPD30	COPD
31	ISG2		COPD-ISG2	COPD31	COPD
32	ISG3		COPD-ISG3	COPD32	COPD
33	ISG4		COPD-ISG4	COPD33	COPD

Appendix C Mobilise-D Ingestion Pipeline