831458 – Trials@Home

Center of Excellence – Remote Decentralised Clinical Trials

WP6 – PROMS

D6.11 Ethics D4, POPD– Requirement No. 4: Updated data management plan

Lead contributor	3 Lygature
Other contributors	1 UMCU 18 SARD

Document History

Version	Date	Description
V1.0	09 Jun 2020	D6.3 First version of Data Management Plan
V2.0	15 Jun 2023	Updated Data Management Plan

Table of Contents

1. Data summary

The main objective of the Trials@Home project is to reshape clinical trial design, conduct and operations, through public-private partnership, by developing and piloting standards, recommendations and tools for the definition and operationalisation of remote decentralised clinical trials (DCTs) in Europe.

Overall, the Data Management Plan (DMP) provides a description of the data management that will be applied in the Trials@Home project including:

- A description of the data repositories, who is able to access the data, and who owns the data.
- The main DMP elements for each of the studies contributing (or sharing data).
- The time period for which data must be stored.
- The standards for data collection, validation and evaluation.
- The possibilities of and conditions for sharing data.
- The implementation of data protection requirements.

The DMP will be updated over the course of the project whenever significant changes arise, such as (but not limited to):

- Addition of new data
- Changes in consortium policies (e.g. new innovation potential, …)
- Changes in consortium composition and external factors (e.g. consortium members and/or associated partners joining or leaving).

The Trials@Home project has identified 2 main types of data that will be collected during the time of the action:

"Clinical Data" means any and all unprocessed information or raw data, including but not limited to any and all serological data, metabolic or genetic data, collected for each human subject involved in a Clinical Trial carried out as part of the Action. Clinical Data to which another Party may have access to in the framework of the Consortium Agreement shall be pseudonymised or anonymised. Clinical Data remain under the sole management, custody and responsibility of the Sponsor and constitute Results under this Consortium Agreement.

"Clinical Trial Metadata" means any and all metadata obtained in a Clinical Trial, that provide information about the design, conduct, performance and compliance of a Clinical Trial, including but not limited to: user preferences (patient, site, sponsor, other stakeholders), interview notes, transcripts or recordings, performance metrics and model relevant KPI's (throughput, recruitment, retention, …), feedback from regulatory authorities and ethical committees, observed incidents and support issues, user satisfaction, and/or cost-benefit results. Individual patient data will be anonymised by the data owner or Sponsor before sharing with any other party.

Work Packages	Types of data
WP1 BEST	• Data on criteria to analyse DCT case studies; • Data on/from DCT case studies available from industries and public/private case studies; • Systematic review data on selection of appropriate DCTs
WP2 TECH	• Systematic review data on suitable DCT technologies; • Quality assessment data of selected technologies
WP3 PILOT	• Data from the feasibility study (dry-run) • Data from the pilot DCT as per final study protocol
WP4 EAGLE	• Data on EU legislation including legal, regulatory, ethical and operational barriers and enablers for DCT in Europe; • Data on technical and regulatory implications of DCTs; • Data on innovative scenarios for clinical trial ecosystems from an ethical and normative point of view; • Stakeholder input on various aspects of elements related to regulatory, legal or ethical questions, analysed using qualitative research methodologies (interviews, focus groups)
WP5 CODE	• Stakeholder interview and podcast data on the consensus on key project deliverables; • Online, and telephone survey data on internal and external stakeholder perceptions, opinions regarding barriers and enablers of DCTs in Europe; • Map of paradigm changes in the relationship between HCPs and Patients; • Information that is provided for the purpose of registering with the website and/or subscribing to the website services (e.g. name, job title, company name, email and phone number)

In summary, the Trials@Home project DMP gives guidance and provides an oversight of general data management, while each study needs to provide specific data management information 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. In cases where the research results are not open access a justification needs to be provided.

2. FAIR data

2.1 Making data findable, including provisions for metadata:

Discoverability:

Audio data: audio recordings will be made of all interviews. These will be transcribed and pseudonymised. All data will be stored in a secure database at the University of Dundee with limited access to research team only. All data will be handled and stored in accordance with local information governance SOPs (Standard Operating Procedures).
Participants will not be individually identified in any publications but nonetheless may be identifiable by role to those with knowledge of the case studies being presented.

Participants do have a right to withdraw their consent to participation at any time. However, rights to access, change or remove their data may be limited if data have already been analysed and/or incorporated into study results. If participants do withdraw from the study they will be informed as to what data has been collected and if it can be removed.
Data collected for this project will not be made available for reuse and will not be used for automatic decision making or profiling.

Datasets that will be made publicly available will be uploaded in open repositories like Zenodo and GitHub, thus making this data both easily discoverable and identifiable from the outside (since they will be given a Digital Object Identifier (DOI). Some real-time sensor data will be shared on the project's websites and portals, where they will also be discoverable.

Datasets that will be only used internally by project partners will be stored either on the project wiki/SharePoint repository, MyProjectPlaza, or in Trials@Home platform instances hosted on a secure cloud. In both cases, the datasets are internally discoverable and identifiable using simple queries with keywords or filters.

Data will be made discoverable in different ways depending on its utility for internal or external stakeholders. Versioning of data, whenever applicable will be applied to all data (incl. documents, questionnaires) created and/or collected. Secondary data will be documented by carefully explaining terms, variable names, codes and abbreviations used.

The data can be organised by naming files using pseudonymous identifiers for study participants (participating in stakeholder interviews as well as those filling in questionnaires) or at different study/test occasions. Because of the interdisciplinary nature of the project, it is important to promote the use of common terminology within the project. The choice of terminology will be driven by the Trials@Home glossary.

Depending on the method of data publication chosen, version number will be made available. The templates created for the project deliverables will include the version numbers and details about the changes made.

Identifiability:
Datasets that will be made publicly available in open repositories will be given a DOI.Interview transcripts will be coded using NVIVO to generate qualitative metadata.

Naming conventions used:
A specific naming convention is used to identify the various Trials@Home datasets: Data_<WPno>_<serial number of dataset>_<dataset title/ID>.

The <WPno> reveals the WP in the context of which this data is collected or generated and processed.

The <serial number of each dataset> is assigned manually in the order of presentation in this deliverable.

The <dataset title> for interview data, medication and sensor big data sources is the same data source identifier (i.e. the DS_ID) as the one used in the context of other deliverables, so as to be consistent in the various project documents. For other types of datasets which are related to user requirements, platform assessment etc. a descriptive dataset title is introduced.

Keywords for searches will be provided in the cases where this is applicable.

For datasets that will be made publicly available in open repositories, versioning will be supported by these repositories. Versioning is also supported by the project wiki MyProjectPlaza.

Metadata creation:
For datasets that will be shared via open repositories, the metadata standards used by these repositories will be used. Metadata for data uploaded at the project wiki is also supported. Same for data stored on the Trials@Home platform repository.

2.2 Making data openly accessible:

The general principle is to keep data closed to external use, except for during phases of the project that it is necessary to open them up, after a clear explanation. Generally, all data is accessible only for Trials@Home members. Further analyses towards copyright material and privacy considerations will determine the publicity of each dataset.

Before the use of any of these data for publication or dissemination purposes beyond the project members, the coordinator should request the approval of the ethics board and advisor. Upon that, the data should undergo an anonymization and de-identification process so that no personal data is included. Then, they can safely be shared out of the project for specific publication or dissemination purposes.

Furthermore, data will need to circulate among some project members. In any case, effort will be made to make some of this data openly available in cooperation with the data owners. At this point, some private datasets will be openly shared after anonymization.

In addition, there is also privately-owned data, usually collected by companies/institutions working with the patients. Such data (e.g. data for the patients' medication or interviews), provided to the project for research purposes, will be used internally by project partners after following the procedure defined in WP1 BEST. Aggregated and anonymised conclusions and abstractions obtained through the analysis of data may be publicly or confidentially shared if this promotes the project's goals. Personal information that may identify individuals will never be shared.

Data to be openly shared will be deposited in open repositories like Zenodo and GitHub or the project's websites. Real-time online but also offline data logs, e.g. of sensor data, may need to be shared via an API to promote integration and building around the Trials@Home platform. Datasets that will be only used internally by project partners will be stored either on the project wiki / MyProjectPlaza or in the project's platform instances.

Generally, standard computer software and no developer skills will be required to access the data e.g. to download, and read them in Excel, CSV formats. However, different methods and software tools will be required to access data depending on the nature of data and means of publication e.g. website, wiki or API. That may include database management systems e.g. SQL, more sophisticated formats like JSON or standardised formats such as LinkedData etc. Where this is applicable, the relevant software and its documentation will be included.

Data to be openly shared will be deposited in open repositories like Zenodo and GitHub. These are widely used repositories adopting standard and simple procedures to allow data sharing by researchers. No need for appropriate arrangements is foreseen. Real-time or data logs may be shared (via an API). Arrangements will be made by the partners participating in the current project. Finally, data shared on the project's website will be arranged by the consortium.

In the case of restricted use, access could be provided either through use of consent and anonymization, or by regulating and restricting access to external users. Access will be provided by the Data Access Committee. The DAC will manage all requests from external parties, evaluate data, maintain inventory and review in-kind contributions.

Machine-readable licenses will be used for the data we plan to make openly available.

For the datasets we plan to share, open access will be granted. For the data that will be used only internally by project partners (which is stored on the project wiki or the project's platform), access control procedures are in place that define access rights and provide secure access with username/password credentials.

2.3 Making data interoperable:

Trials@Home integrates real-world data coming from diverse sources, like sensors, patients' medicine and interviews. To be able to easily integrate, analyse and share these diverse types of data, techniques for data harmonisation are being implemented and significant effort has been made by project partners to ensure data interoperability.

Techniques for data harmonisation are being implemented and well-known standards will be adopted to ensure interoperability. Data harmonisation aims to align data from heterogeneous

sources into a coherent and unambiguous set. To deal with various naming and structural differences among diverse data sources, we will adopt specific data representation models for different types of data: medication, interviews and sensor data. In addition, to facilitate the exchange of information and sharing of data, in this project we rely on accepted and widely used formats and standards. More information on these standards can be found below.

To facilitate the exchange of information and sharing of data, in Trials@Home we rely on accepted and widely used formats and standards. Examples may include:

- Date and Time information is compatible with the ISO 86012 representation.
- ISO IDMP for medicinal products
- HL7 for medical data exchange
- The Semantic Sensor Network Ontology for sensor data (https://www.w3.org/TR/vocab-ssn/)

It will be decided on a case-by-case basis whether we will provide mappings to more commonly used ontologies when using uncommon or project specific ontologies or vocabularies.

2.4 Increase data re-use:

Data needed to validate the results presented in scientific publications will become accessible via Controlled Access directly after publication. Datasets that will be made publicly available will be uploaded in open repositories like Zenodo and GitHub, thus making this data both easily re-usable by outside parties (since they will be given a DOI). Some real-time sensor data will be shared on the project's website and portals, ensuring re-usability.

The accessibility to other data which will not (yet) be published will be determined in a later phase of the project. Trials@Home data will be accessible for at least five years after the end of the project. Efforts will be made to prolong this period to ten years.

An important aspect of data reusability is the quality of the produced data. To ensure the quality of the datasets, several quality assurance processes are in place, each being the responsibility of the Work Package generating the data.

3. Allocation of resources

Open Access costs will be covered by project budget of the partner organisation for which the corresponding author is originating from. Cost for storage, collection and analysis of collected data will be covered by partner institutions responsible for the relevant deliverables.

Data Management (and all associated costs) for the pan EU-pilot study is the responsibility of WP2 TECH, who set up a database and developed a detailed data management plan for RADIAL, the pan EU- pilot (available as Deliverable 2.5, see DoA (Description of Action), task 2.4).

4. Data security

Data security is the responsibility of the various partners generating data and should be monitored by each institution's DPO (Data Protection Officer).

Data transfer between partners

Trials@Home researchers commit to the highest standards of data security and protection to preserve the interests of the study participants. Data security when sending patient data between Trials@Home partners is ensured by using the facilities offered by the data generating institute. Data Protection Officers of the institutes will be consulted to find the appropriate solution, which in most cases will be an encrypted data transfer requiring licenses, such as e.g. ShareFile.

5. Ethical aspects

Patients will be informed of the study and study-related risks, asked to read the information sheet for the patient, ask questions when needed, and to sign informed consent before any study-related procedures take place. The informed consent template is based on local legislation and ethical committee prescriptions. Future re-use of (patient) data is included in the informed consent.

After inclusion for screening, the patient's name will be converted to a unique encoded patient ID and all analyses will be performed based on this unique ID. During the complete course of the study, all patients will be identified by their patient study number. The pseudonymisation key (data for personal identification) will be kept separate from the eCRF or other patient data, and will only be accessible to dedicated consortium members: PI, project manager and neurologists at the various clinical sites.

The ethical requirements described in the clinical protocol set specific requirements for the anonymization of data and protection of personal data of patients, as does the GDPR legislation. These requirements will be strictly followed, which prevents the open sharing of patient-level data. For this reason, as described earlier, clinical data will only become accessible beyond the Trials@Home consortium under Controlled Access. The Data Transfer Agreement will ensure that future re-use of the data is within the restrictions of the Informed Consent.

6. Other

The DMP is not a static document but will be updated during the progress of the project. Each updated version of the DMP will contain a more detailed and updated description of the datasets.

A detailed data management plan, specific for the pan EU-pilot, is developed by WP2 TECH (Deliverable D2.5). This is a confidential document.

A final version of the DMP is planned as Deliverable D6.8. This will contain more information about how the project will go about open data publishing.