RealHOPE

innovative medicines initiative

IMI2 Project ID - RealHOPE

Real World Handling of Protein Drugs Exploration, Evaluation and Education

WP5 - PAGE

D5.5 – Interim Version of Data Management Plan

Lead contributor	University of Dundee (4)
Other contributors	RISE (3), TEVA (24), LU (1), CPI (2), SANOFI (23)

Document History

Version	Date	Description
V1.0	01/11/2021	First version of Data Management Plan
V2.0	27/06/2023	Interim version of Data Management Plan

efpia

Contents

1. Data Summary

The main objective of the RealHOPE project is to identify key challenges in the post-production handling of protein drugs products and verify if these can represent a threat for products' stability and efficacy. Following that, the consortium will focus on the generation and implementation of measures to mitigate and improve the handling of protein drug products. Training and education for stakeholders, revision of current guidelines (by establishing a dialogue with the main regulatory bodies) and the use of tools to support the preparation and handling of these products in clinical and non-clinical settings will be implemented. The RealHOPE project will involve four general types of data:

1) "Interview data" collected from qualitative interviews, focus groups and questionnaires among the stakeholders involved in the day-to-day handling of protein drugs products (e.g., patients, hospital pharmacists, nurses, distributors…).

2) Data from smart labels. They will include a set of environmental tests and parameters, with associated time stamps. The data collected via the labels will be pseudonymised as it will have a study participant ID associated with it from the start. All data stored on the label at the end of the monitoring period will be downloaded onto an intermediate companion device – a study mobile phone – and will then be transferred to a secure server within University of Dundee. The raw data will be JSON files which will be converted into .CSV files for analysis).

3) Data collected from the "In-lab simulation studies". They will consist mainly of analytical data focussing on the physical and chemical quality of the drug product. This will concern data from compendial and purity methods specific to the molecule used in the study.

4) Data collected from patients taking part in the Investigational Study in the form of patient diaries/questionnaires recording events during the course of the monitoring of handling of their protein drugs from collection/delivery to administration. This data will be identified by a participant number (pseudonymised) and stored in the University of Dundee. Patient-identifiable data from the study e.g. name, date of birth, contact details of participants will be held securely in the University of Dundee and will not be shared.
The key linking the study participant ID and the identifiable data will be stored on the study secure server and will only be accessible by approved University of Dundee study team members.

The data will be collected by the different organisations involved in the project and remain under the management, custody and responsibility of the specific organisations.

Overall, the data that will be collected from existing data, generated *de novo*, and analysed (metadata) during the project have been identified as follows:

Work Packages	**Types of data**	**Format**
WP1	Data gathered from literature review	Articles, grey literature, websites, blogs
	Ethics approval supporting documents	Reports
	Data from qualitative interviews/focus groups/round tables, surveys with stakeholders	Recordings, transcriptions
	Raw sensor data from smart labels (i.e., temperature, light exposure, shock…).	JSON files, .csv files
	Data from patient investigational study in the form of participant diaries/questionnaires	Reports
	Patient identifiable data for participants in investigational study	Text
WP2	Data from laboratory simulation analyses of protein drugs	Analyses results, laboratory journals (paper and/or electronic)
	Data from method development	Instrumental outputs, laboratory journals (paper and/or electronic)
WP3	Data from the development of new technologies for stress factors mitigation	Reports, laboratory journals (paper and/or electronic)

	Data from guidelines produced as a result of the technologies developed	Reports
WP4	Data generated from stakeholder interviews and surveys	Audio recordings, transcriptions.
	Guidelines, training and educational material generated as a result of the data generated in WP1, WP3 and from stakeholders' interviews	Videos, brochures, posters, reports.
	Platform development	Web construction
	Data from regulatory and reporting activities	Reports
WP5	Management Plan	Report
	Dissemination and exploitation plan	Report
	Data generated as a result of communication and interaction with IMI	Report
	Budget data	Report/other

The RealHOPE Description of Action (DoA) refers to a Data Management Plan (DMP) as a deliverable D5.2 as part of WP5 PAGE. The interim version of the DMP is deliverable D5.5. The DMP will be updated again at the end of the project (Final DMP D5.6).

The Data Management Plan (DMP) provides a description of the data management life cycle that will be applied in the RealHOPE project.

This will include:

A description of the data repositories, who is able to access the data, and who owns the 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 handled following an adaptive approach and therefore 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 composition and external factors (e.g., consortium members and/or associated partners joining or leaving)

2. FAIR Data

2.1 Making data findable

Where possible, data will be findable, accessible, interoperable and reusable (FAIR) according to the Guidelines on FAIR Data Management in Horizon 2020.

Discoverability

Audio data: qualitative interviews with stakeholders (English, Swedish and Italian speakers) will be audio recorded. These will be translated (into English, when needed) transcribed verbatim and pseudonymised. All data will be stored on a secure server at the associated university, with access limited to the research team only. All data will be handled and stored in accordance with local information governance Standard Operating Procedures (SOPs). The data will be organised by naming files using pseudonymous identifiers for study participants (e.g., those participating in stakeholder interviews as well as those filling in questionnaires) to maximise anonymity of the participants.

Transcriptions of audio files recorded in the UK are made using a University of Dundee approved service supplier – TP Transcriptions. Audio files are password protected and transferred directly to a secure company server. The transcriptions are returned to the study team at the University of Dundee via the same password protected, secure route.
Audio files recorded in Sweden are transcribed and translated within the study team.

Participants do have the right to withdraw their consent to participation at any time, however, rights to access, change or removal of data may be limited once it has been analysed and/or incorporated into study results. If participants decide to withdraw, they will be informed as to what data has been collected and if it can be removed. Data collected for this study will not be made available for re-use outside of the study team, and will not be used for automatic decision making or profiling. At the end of the study (June 2025) all audio recordings will be destroyed. Pseudonymised transcripts will be archived with study data for 10 years. The pseudonymisation key will be retained with access restricted to approved study team and Sponsor personnel.

"Smart Label" studies: Raw sensor data will be collected (dependant on the study or configuration of the label – i.e. temperature, light exposure, shock, etc) over the lifetime of the label battery. The data gathered will be transferred from the label onto an intermediate device – a study mobile phone.

The psuedonymised smart label data from the patient investigational study will be transferred to a predefined location on a University of Dundee secure server in a predefined format (JSON files (raw data) and .csv files for analysis). Data handling, accessibility and organisation are yet to be defined by the consortium.

The data generated from the samples in the in-lab simulation studies will be coming from multiple analytical instruments. In addition, there might also be data generated by manual recording, for example visual appearance of sample. Nonetheless, the majority of the data generated in this stage of the program will concern data obtained from analytical methods that assess the physical and chemical stability of the molecule of concern. The electronic data in the form of pdf or csv files will

be stored on protected servers of the corresponding entity where the measurement took place. The data retention and back up policy of the relevant entity will be applied for the collected data.

Datasets to be only internally used and discoverable by project partners will be stored either on the project SharePoint or on the RealHOPE platform where they will be hosted on a secure cloud. Datasets will be internally discoverable and identifiable using simple queries with keywords or filters.

Any datasets that are made publicly available will be hosted in open access repositories (still to be decided) and discoverable through an assigned Digital Object Identifier (DOI).

Versioning of data, whenever applicable, will be applied to all data (including documents, questionnaires) created and/or collected. Secondary data will be documented by carefully explaining terms, variable name, codes and abbreviations used.

Project deliverables will be drafted and finalised using version control (with "reasons for revision" documented) templates.

Identifiability

Interview transcripts will be coded to generate qualitative analysis. Most likely NVIVO, Python3 and Matlab software will be used to analyse the data. Identifiable data will be stored on secure servers in Lund University (Swedish data) and University of Dundee (UK data).

The following naming conventions for final datasets generated will generally be used to easily identify the different RealHOPE datasets:

<Date>_<RHWPno>_<serial number of dataset>_<X dataset title/ID>_<version no>

- **<Date>** related to the document's version (format DDMMYY).
- **<RHWPno>** RH (RealHOPE) for the WP of which this data is collected or generated and processed.
- **<serial number of each dataset>** assigned manually in the order of presentation for the different deliverables.
- **<dataset title> X:** a unique code before the dataset title will be created and used for the specific dataset (e.g., "I" for one-to-one-interviews, "F" for focus group interviews…)

  max 50 characters (with spaces).
- **<version no>** to match the one in the document

Metadata generation

Metadata with public access will be assigned with a DOI and deposited in open repositories according to their standards. The same process will be followed for the data deposited and shared on the organisations repositories.

ORCID Registration

Authors of documents with open access will register at ORCID and will use the personal persistent author identifier for all the available publications.

The persistent identifiers will be referenced in the research output and the DOI metadata.

2.2 Making data openly accessible

In general, data will be kept closed to external use, except for data generated for public purposes or for phases of the project where it will be necessary to open them up, after a specific and clear reason. Data will be accessible only for RealHOPE consortium's members and some will be restricted to access by particular members or groups where this is necessary.

Before any data sharing occurs, data will undergo a pseudonymisation and de-identification process. No personal data will be included in the process of making data openly accessible. Once this process is completed, it will be possible to share the data safely for publication, dissemination and/or education purposes. Anonymised conclusions and abstraction obtained through the analysis of data may be publicly or confidentially shared to promote the objectives and goals of the project. Personal information that may identify individuals will never be shared as part of this process.

Some data will be shared among project members of different institutes collaborating on the same tasks. Such data will be used internally for research purposes and will be shared according to the SOPs and procedures defined by the specific institutes and in accordance with the General Data Protection Regulation (GDPR) European Union (EU) 2016/679 (2018).

Data to be openly shared will be deposited in open repositories (to be established) or the RealHOPE project website.

Generally, standard computer software and no specific "Information Technology" (IT) skill will be required to access the data. If a specific software will be required to access a certain dataset, when applicable and possible, the relevant software will be included.

Data links shared on the RealHOPE project website will be approved and arranged by the members of the consortium.

2.3 Making data interoperable

RealHOPE will combine data coming from diverse sources such as interviews, smart labels and laboratory instrumentations. In order to promote interoperability and future-proofing of the data, we will use standardised data storage formats. Recordings of the focus group interview will be carried out using digital format (M4A audio). Written documents will be in MS Word (.docx) and .pdf formats. Smart labels data, databases and statistical data will likely be in .csv format. Pictures will be stored in .png, .svg or .jpg formats. Videos, if any, will be stored in MP4 and M3U formats. There may be occasional circumstances when other formats are required.

2.4 Increase data re-use

To guarantee the quality of the datasets generated, assurance processes will be in place and will be controlled by the work package generating the datasets. In general, and where possible, efforts will

be made by the work packages to promote data re-use by organising and annotating the completed data sets for easy identification and analysis.

3. Allocation of Resources

The project budget of each partner in the consortium will cover their own costs for Open Access publications. Costs associated with the storage, collection and analysis and any data collected will be covered by the partner organisations responsible for the specific activities/deliverables.

4. Data Security

Consortium Member Responsibilities

The Data Protection Officer (DPO) of each organisation involved in generation of data will be responsible for data security. Each consortium member must confirm that contact details of the Data Protection Officer (DPO) are made available to all data subjects involved in the research. For the members of the consortium not required to appoint a DPO under the General Data Protection Regulation, a detailed data protection policy for the project must be maintained and must be provided to IMI JU upon request. A description of any technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants must be provided.

Consortium members must ensure that any special derogations pertaining to the rights of data subjects or the processing of personal data have been established under the national legislation of the country where the research takes place. They should provide a declaration of compliance with respective national legal framework(s).

Details of the observation/monitoring activities (such as shadowing of healthcare personnel and monitoring the handling of protein drugs using smart labels attached to drug containers) must be provided. It will be explained to participants whether or not these activities will involve personal data collection/processing.

If it is not contained within the consent process, detailed information on the specific data protection informed consent procedures in regard to data processing must be provided prior to the start of the research activity that raises an ethical issue. A description of the anonymisation/pseudonymisation techniques that will implemented must be provided where it is used.

Data transfer between partners

- Qualitative interviews activities:

RealHOPE researchers will adhere to the highest standards of data security and protection to preserve the interests of the study participants.

Data transfer (after pseudonymisation and de-identification process) between EU countries and UK (and vice-versa) will adhere to GDPR EU (General Data Protection Regulation) 2016/679 (2018) rules.

In the event of a data transfer (after pseudonymisation and de-identification process) from the EU/UK to a non-EU country or international organisation (USA/Israel), the process will be performed and documented in accordance with the General Data Protection Regulation 2016/679 (GDPR) rules.

In the event of a data transfer (after pseudonymisation and de-identification process) from a non-EU country (USA/Israel) to the EU/UK partners, the process must comply with the laws of the country in which the data was collected.

The Data Protection Officer for each organisation involved in data transfer will be consulted to find the most appropriate solution.

- Other activities:

The applicants will explain if any research activities to be performed in a non-EU country will result in transfers of data or materials to/from non-EU countries. If this is the case, the concerned consortium members will provide details on the materials (e.g., smart labels) which will be imported/exported into/from the EU and confirm that the adequate import/export authorisations required by national/EU legislation have been obtained prior to the start of the research activity.

Archiving and deletion of data

Archiving and deletion of data will be safely implemented according to the SOP and/or specific policies of the organisations involved in the project, however, RealHOPE data will be accessible for at least five years after the end of the project.

5. Ethical Aspects

Qualitative Interview Data

Ethics approval will be obtained with the local authority for the organisations involved in the qualitative interview activities (WP1 and WP4) prior to initiating the interviews. SOPs and procedures related to the human-ethics process of data collection will be in place and followed by the researchers performing the interviews. Participants will be approached by email and provided with a participant information sheet and consent form. Prior to giving their consent for the interview, participants will have the opportunity to ask questions related to the project/qualitative interviews to the research team that will lead the interviews. The interview preamble will include an explanation that participation is voluntary and that participants have the right to stop the interview at any time. Deception will not take place when carrying out the interviews. Participants taking part in qualitative interviews, questionnaires and focus groups will be identified by a unique ID which will be pseudonymised and will only be accessible to specific members of the research group of the organisations performing the interviews and involved in the qualitative interview activities.

Patient Investigational Study

Ethics approval will be obtained within the UK for the investigational patient study in WP1 SHAPE. Data collected will be from smart labels attached to their prescribed protein drugs and also directly from participants via patient diaries/questionnaires. All participants will be identified by a study ID number and all identifiable information will be stored separately from the study data within a secure

server in the University of Dundee. Participants will have consented to their data being stored and their non-identifiable data e.g. data recorded by smart labels, diaries/questionnaires being shared with consortium partners as needed for extraction and analysis. No patient identifiable information will be shared outside the study team of the University of Dundee.

5.1 Unexpected findings

It is possible that during the interviews or within the participant diaries unexpected information will be disclosed by the participant. This information could be related to criminal activity, unintentional or intentional mishandling of the medications (that could potentially cause immediate harm) or sensitive information related to the participant's health. This information will be dealt with according to the law and regulation of the country the participant resides in and considering the professional code of conduct that the interviewer or researcher must adhere to.

6. Other

The DMP will be a live document and will follow an adaptive approach. This means that the document will be updated during the course of the project whenever needed. Each new version will contain more precise information related to the collection, sharing and processing of data if these will be different from the current used version. This is the interim version (D5.5) correct at month 24 (June 2023). The final version of the DMP will be provided at month 48 (D5.6).

7. List of abbreviations

DMP Data Management Plan

DOI Digital Object Identifier

DS Data Source

EU European Union

GDPR General Data Protection Regulation

IT Information Technology

SOP Standard Operating Procedure

WP Work Package