

Data Management Plan (Model)**Data Management Plan (DMP) for clinical research projects****A model for submissions to the Swiss National Science Foundation (FNS) and/or to Ethics Committees****Important Notes:**

This document is made available by UIC (*Unité d'Investigation Clinique*), a unit which is part of CRC (*Centre de Recherche Clinique / CTU*) at HUG (*Hôpitaux Universitaires de Genève*) and the Faculty of Medicine UniGE (Geneva University). It aims to guide clinical researchers through the submission process of their Data Management Plan (DMP) to FNS (or to Ethics Committees). It also aims to provide clinical researchers with the technical description of how their research data will be managed by UIC, assuming they would have contracted or plan to contract UIC to be in charge of the data management of their submitted project.

UIC is certified ISO 9001/2015, and the unit guarantees best practices in the field of clinical data management. It is however necessary to emphasize that submission of this document is made under the responsibility of the grant applier, commonly designated as *sponsor* of the study. In order to engage UIC responsibility regarding any topic listed below, the sponsor ought to contract UIC for the management of his/her clinical study data.

Following an agreement between the sponsor and UIC, all technical responsibilities described below (*cf.* appropriate colors code) are ensured by the unit and can therefore be submitted without further modifications by the sponsor. Project-specific topics should be reviewed, depicted, altered or removed by the sponsor. This will depend on the type of services the sponsor may or may not wish to contract with UIC or with any other service provider, or is able to provide by himself.

It is therefore the duty of the sponsor to adjust this Data Management Plan Model and submit it, under his responsibility, to FNS (or to Ethics Committees).

It is also important to bear in mind that for his/her initial submission to FNS (SNF), the sponsor is only requested to describe his/her intentions regarding the data management plan. This means that he/she is fully allowed to amend his DMP subsequently to reflect any change throughout the project conduct. Only the final DMP version at the time of project closure will be binding.

Unité d'Investigation Clinique (UIC)

Data Management Plan (Model)

Data Management Plan – content of the mySNF form

Colors code

- specific to data management by UIC/CRC
- choose between propositions
- optional
- project-specific (remove/add what is applicable for your project)

1. Data Collection and documentation

1.1 What data will you collect, observe, generate or reuse ?

Data management will be performed by UIC (Unité d'Investigation Clinique), a unit which is part of CRC (Centre de Recherche Clinique / CTU) at HUG (Hôpitaux Universitaires de Genève) and the Faculty of Medicine UniGE (Geneva University). UIC is certified ISO 9001/2015, and the unit guarantees best practices in the field of clinical data management.

Data are physically stored in a [Oracle ver.12c/19c *for secuTrial®* | *mysql ver.7.x for REDCap™*] RDBMS (Relational Database Management System) using a dedicated CDMS software (Clinical Database Management System) [secuTrial® | REDCap™].

All research subjects will be [pseudonymised / anonymised]. Data that are collected for each research subject constitute the subject CRF (Case Reports Form), which cover the following areas/topics:

- Inclusion criteria and consent
- Demographics
- Medical history
- Concomitant medication
- Laboratory analysis
- Medical events
- Hospitalisation and interventions
- Surveys and questionnaires
- Adverse events
- Binary files (images, radiography, scanners, etc.)
- ...
- Randomisation process of subjects will be performed [automatically following his/her inclusion / {or describe any non-automatic process}]. (Double-)blinding will be

Unité d'Investigation Clinique (UIC)

Data Management Plan (Model)

applied...

1.2 How will the data be collected, observed or generated ?

A dedicated CDMS project will be created to reproduce the CRF electronically: eCRF (electronic Case Report Forms). This eCRF will reflect the visit plan and all subject related data to be collected, as described by the study protocol. CRF data are of different data-types, mainly numeric values, encoded values, text descriptions, etc. Coding, catalogs and multi-choice values will be preferably used as opposed to free text values. Automatic score calculations will be performed whenever appropriate.

The eCRF will ensure high data quality control by applying logical and managerial controls over data capture in order to ensure consistency, completeness and coherence of data.

A more exhaustive and detailed Data Management Plan will be generated and applied during the whole data management process (until data are archived), assuming the sponsor have obtained the grant and contracted UIC.

The investigation team will enter the data interactively in the CDMS by means of its interactive graphical interface {some subjects' surveys will be directly filled up by the subjects themselves}. Data will be edited and viewed by the investigation personnel based on the role of each participant. Validation/review will be performed in parallel by appropriate people designated by the sponsor {monitors, sponsor, coordinators, data managers, etc.}. Up-to-date graphical and text reports, as well as descriptive statistics, will be made available to monitor data in real time. An elaborated electronic process to comment, query and claim correction actions on data via reviewing will reinforce data quality throughout the study conduct and termination.

Whenever appropriate, use of controlled vocabulary and common international catalogs (ATC/DCI, MedDRA, NCBI Taxonomy, etc.) will be used.

Data and metadata will be exported in due time in plain text (CSV, SPSS) or SAS formats. These exports might be post-processed programmatically by UIC to facilitate data analysis and data visualisation.

All modifications to the eCRF during the study conduct will be tracked and dealt with consequently (version numbering). Data will be electronically and systematically revalidated after each major change, and corrections/completeness applied whenever necessary.

All system log-ins, interventions, modifications and form status changes will be thoroughly

Unité d'Investigation Clinique (UIC)

Data Management Plan (Model)

recorded in an Audit Trail log system. Deviations to the protocol will be also tracked and resolution status indicated.

1.3 What documentation and metadata will you provide with the data ?

Raw data will be provided with its full description, data-types and encoding (comprehensive Code Book). Metadata regarding database implementation, audit trail, data status, completeness, modifications and intervention dates will be available for each single value. A complete description of the whole project and its technical configuration, as well as a complete version history of structure alteration/modification to the database implementation will be made available upon request. All deviations will similarly be logged in.

Any post-processing of data will be fully described by the [description of all transformation operations | the software code applied to the data].

Metadata will be provided as much as possible in non-proprietary formats (text, CSV, XML, PDF). However, some material may remain proprietary due to technical restrictions.

Relevant additional data material not present within the main CDMS database will, whenever necessary, be attached to the raw data, and will be fully described.

Depending on regulations, we will consider applying CDISC (Clinical Data Interchange Standards Consortium, cf. <https://www.cdisc.org/>) standards and data transformations in due: time {indicate if sponsor has specifically contracted UIC, or any other service provider, for CDISC compliance and data transformation}.

2. Ethical, legal and security issued

2.1 How will ethical issues be addressed and handled ?

Even though research subjects will be [pseudonymised / anonymised], consent of research subjects is mandatory in order to enroll them in the study and collect their data {indicate if – exceptionally - consents will not be collected}. No personal data or data that may easily identify subjects will be provided, with respect to the Swiss law on human research (Federal Act on Research involving Human Beings [HRA]) and its applicable ordinance: HRO/ ClinO/ClinO-MD.

Unité d'Investigation Clinique (UIC)

Data Management Plan (Model)

Approval from the competent authority (e.g. Swissmedic) and from an appropriately constituted Competent Ethics Committee (CEC, e.g. CCER) is mandatory before conducting the study.

Only eligible subjects will be enrolled. Any research subject is entitled to withdraw from the study at any stage. He/she is also entitled to require a complete removal of his/her data from the study database upon request.

Data will remain accessible for authorised people throughout the study conduct, and until the final publication of results (if applicable).

2.2 How will data access and security be managed ?

Physical access to the data centers is logged and limited to authorised personnel using badge authentication. On a regular basis, vulnerability testing is performed to reduce potential exposure.

Remote access to servers is limited to authorised personnel. Connections to servers are encrypted using SSH. System logs are stored in a dedicated centralised system for audit purposes. The internal HUG network is protected by multiple firewalls, proxy, reverse-proxy and anti-virus solutions.

Web servers operate under SSL (HTTPS) certifications, ensuring Web connections are encrypted and secure.

At the CDMS level, only people part of the investigation team, the sponsor team, the affiliated reviewers or auditors, as well as inspection authorities (Swissmedic) are given access to data. Personal accounts are granted individually for each person. Identification is made by a personnel ID and a password. Failure to provide the correct password after a limited number of attempts automatically deactivates the faulty account (protection against non-authorized attacks).

Only institutional e-mail addresses will be accepted for any communication of sensitive data regarding account creation and management.

Pen-Tests (simulation of malware attacks) are regularly performed, and measures taken whenever necessary.

Data transfer: Exports of all or any kind of partial data will be systematically password encrypted before being transferred. Use of hashing encoding will ensure that no data

Unité d'Investigation Clinique (UIC)

Data Management Plan (Model)

alteration may have occurred during the transfer.

2.3 How will you handle copyright and Intellectual Property Right issues ?

All data is the property of the sponsor institution (HUG), and is under the direct responsibility of the sponsor.

Intellectual Property Rights are defined as follows: *{define rules regarding intellectual property and publications, consortia, co-authorship, etc.}* – see <https://creativecommons.org/licenses/>

{example of answer 1, the most appropriate} The research is not expected to lead to patents. IPR issues will be dealt with in line with the institutional policy. As the *{anonymised data / metadata}* is not subjected to a contract and will not be patented, it will be released as open data under Creative Commons CC0/CC BY license.

{example of answer 2, extra-institution collaboration} This project is being carried out in collaboration with an industrial partner. The intellectual property rights are set out in the collaboration agreement. The intellectual property generated from this project will be fully exploited with help from the institutional Technology Transfer Office. The aim is to patent the final procedure and then publish the work in a research journal and to publish the supporting *{anonymised data / metadata}* under an open Creative Commons Attribution (CC BY) license.

{example of answer 3, the less restrictive, not suitable in general for clinical research} Data is suitable for sharing. They are observational data (hence unique) and could be used for other analyses or for comparison among many things. Reuse opportunities are vast. For this reason, we aim to allow the widest reuse of our *{anonymised data / metadata}* and will release them under Creative Commons CC0 or CC BY.

3. Data storage and preservation

3.1 How will your data be stored and backed-up during the research ?

All data and applications are physically stored in dedicated data centers on HUG premises. The physical hardware consists of enterprise-grade servers, networking, and storage solutions from tier 1 vendors and trustworthy and stable GNU/GPL solutions.

Data are physically stored in a [Oracle ver.11g *{for secuTrial®}* | MySQL ver.14 *{for REDCap™}*]

Unité d'Investigation Clinique (UIC)

Data Management Plan (Model)

RDBMS (Relational Database Management System) using a dedicated CDMS software (Clinical Database Management System) [secuTrial® | REDCap™]. This CDMS is a central Web-based system consolidating all CRF related data, whether the project is mono or multi centric.

All applications are hosted on certified virtual servers running Red Hat Linux Enterprise in a VMWare ESX environment. There are several dedicated servers for each system, ensuring separation between the testing / pre-production environment and the production environment. System updates are applied only after validation in the pre-production environment.

{only for secuTrial®} For security reasons, the application tiers and the database management systems run independently in separate servers.

HUG infrastructure is under the responsibility of DSI (Direction des Systèmes d'Information at HUG). All exploitation, monitoring and backups operations are performed by DSI, in accordance with UIC policies.

All systems and applications are continuously monitored. Appropriate measures are automatically taken whenever an alert is issued.

Backups operations are performed by the DSI service at HUG, in accordance with UIC policies.

Frequent backups are performed using the best enterprise backup solutions at HUG, and are physically stored in a fire-proof safe. Backup strategy comprises an optimised hourly, daily, monthly and yearly retention plan:

Backup plans:

{only for secuTrial®}:

- Server backups (DSI): 1x differential per day (24 days retention), 1x full every 24 days (retention 12 months), annual (preserved infinitely)
- Database backups (DSI): every day, with 2 full backups twice per week (double copies, 2 months retention for Sunday backups, and 14 days for the others), log backups every hour (2 months retention)

{only for REDCap®}:

- Server backups (DSI): 1x differential per day (14 days retention, with mirroring), 1x full every month (retention 12 months), annual (preserved infinitely)
- Database backups (UIC): hourly, (retained 24 hours), daily (retained 2 months), monthly (preserved infinitely)

Unité d'Investigation Clinique (UIC)

Data Management Plan (Model)

3.2 What is your data preservation plan ?

At the end of the project, the entire database will be archived in a reusable format. Archives encompass all raw data, meta-data, transformed data, transformation operations, deviations, version history, and audit trails. Redeployment of the entire database is therefore possible whenever needed.

The comprehensive archive will remain propriety of the sponsor and will be preserved during a minimum period of [20] years. UIC will also ensure an electronic copy of the archive will be stored within the DSI infrastructure for a theoretical infinite period of time.

Data will be provided to authorised third parties as much as possible in non-proprietary formats (text, CSV, XML, PDF). However, some material will remain proprietary due to technical restrictions.

All Additional data material not present within the main CDMS database will be linked with the raw data and fully described.

Depending on regulations, we will consider applying CDISC (Clinical Data Interchange Standards Consortium, cf. <https://www.cdisc.org/>) standards and data transformations in due time: {indicate if sponsor has specifically contracted UIC, or any other service provider, for CDISC compliance and data transformation}.

4.

Data sharing and reuse

4.1 How and where will the data be shared ?

Individual research subjects' data cannot legally nor ethically be made available to non authorised people (HRA). According to the research subjects' informed consent, only the sponsor, the investigation team, reviewers, auditors and inspection authorities are entitled to access the raw data.

Note for sponsor: at the present time the issue has not yet been fully addressed at HUG, be it ethically or politically. This may evolve in the next few months/years, and you are permitted to

Unité d'Investigation Clinique (UIC)

Data Management Plan (Model)

review this section at the right time (before project closure).

Indicate if you will register your project in any repository outside of HUG, as well as any metadata repository: e.g. Yareta, cf. §4.4. Indicate how you will publish your data and what kind of material will be made available through your publication for third party review or reuse, e.g. descriptive statistics, analysis, data transformation processes, etc.}

Very IMPORTANT: *{for sponsor: In order to help financially researchers to share their research data, FNS is willing to finance research projects with an amount up to 10'000.- CHF. This is intended to cover/participate on the cost of preserving and sharing your data}*

Therefore, DO EXPLICITLY ASK in your budget an amount dedicated to data sharing. This is meant to cover the extra work you will carry to annotate and structure your data to make it sharable, as well as the cost of submitting your data to a FAIR Data repository, which is usually a non-free service.

4.2 Are there any necessary limitations to protect sensitive data ?

Individual research subjects' data cannot legally nor ethically be made available to non authorised people (HRA). Only the sponsor, the investigation team, reviewers, auditors and inspection authorities are entitled to access such data.

No personal data or data that may easily identify subjects will be provided, with respect to the Swiss law on human research (Federal Act on Research involving Human Beings (HRA)) and its applicable ordinance HRO/ ClinO/ClinO-MD.

4.3 I will choose digital repositories that are conform to the FAIR Data Principles

In answer to the statement: "The SNSF requires that repositories are conform to the FAIR Data Principles (Section 5 of the [guidelines for researchers](#), SNSF's explanation of the [FAIR Data Principles](#)). If there are no repositories complying with these requirements in your research field, please deposit a copy of your data on a generic platform (see [examples](#)). **If no data can be shared, this is a statement of principles.**"

{for sponsor} This is to indicate if you will submit your protocol description, META-data of your study, descriptive analysis, or statistical analysis results, to a FAIR Data compliant repository. At the time being Yareta has been approved by UniGe & HUG =>

CHECK 'YES' as a statement of principles.

Unité d'Investigation Clinique (UIC)

Data Management Plan (Model)

4.4 I will choose digital repositories maintained by a non-profit organisation

{for sponsor: indicate if you will submit your protocol description, META-data of your study, descriptive analysis, or statistical analysis results, to a FAIR Data compliant repository, otherwise indicate that at the present time no appropriate solution – politically and technically speaking - is yet considered}

Recommended FAIR repository: **Yareta** (<https://yareta.unige.ch/>).

Yareta is an academic repository compliant with the FAIR principles. It has been developed and is hosted by the University of Geneva. The repository has been approved by swissuniversities, the University of Geneva and HUG.

For more information, check the following links:

- FNS – Explanation of the FAIR DATA Principles: http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf
- The FAIR DATA Principles: <https://www.force11.org/group/fairgroup/fairprinciples>

As a UniGE researcher, we highly recommend you to visit and thoroughly check the following links:

- **Yareta** (<https://yareta.unige.ch/>).
- **UniGE - Données de Recherche**: <https://www.unige.ch/researchdata/> ('Rédiger un plan', 'Stocker vos données', 'Partager vos données')
- **DLCM**: <https://dlcm.ch/>: "The DLCM project involves eight partners institutions (FNS, UniGE, EPFL, ETHZ, UniBS, UniZH, UniL, etc.). It was initiated on September 1st, 2015 following a two-year period of maturation. Given that each partner institution has already major issues in data management to resolve, the involved national funding program provided us with the just-in-time opportunity to gather our respective forces and to grapple with this important and complex topic so as to serve the whole Swiss research community"