Research Data Management

Standard operating procedure

SOP-id	RDM001
version	4.0
date effective	??-??, 2019
Authors	R.A. Scholte, Head Data Management, Clinical Research Unit, location AMC W.G.C. Kraan, Head Data Management, Data Management, location VUmc
Advisory board	S.D. Olabarriaga, Leader e-science research, location AMC J.B.M. Inge, Data Protection Officer, location AMC J.G.M. Stolwijk, Manager Clinical Research Bureau, location VUmc
Authorized by	J.A. Romijn, C.H. Polman, CEO
signature	
date	
Released by	M.J.A.P. Daemen, Chair Amsterdam Research Board
signature	
date	
Distribution	Only electronic version available
Mode of distribution*	– *Kwadraet* and *Kwaliteitsnet* – Intranet pages CRU and Datamanagement – Hard copy filed by CRU and Datamanagement

*Only the on-line electronic version is considered to be up to date. Off-line versions and paper-based versions are not determinative.

Definitions and abbreviations

Audit trail	A record showing who has accessed a computer system and what operations he or she has performed at a specific time point.
CCMO	*Centrale Commissie Mensgebonden Onderzoek, Central Committee on Research Involving Human Subjects.*
CMP	*Centraal Meldpunt Persoonsregistraties, Central registration of personal data registers.*
CRF	Case Record Form; either on paper or electronic:
eCRF	Electronic Case Record Form: the electronic equivalent of a paper Case Record Form (pCRF). By means of an eCRF, data can be entered and stored in a database.
pCRF	Paper Case Record Form: a set of paper forms meant for the collection of study data of a study participant. In this SOP, the term pCRF includes also paper questionnaires.
CRU	Clinical Research Unit, location AMC.
CRB	Clinical Research Bureau, location VUmc.
CTMM-TraIT	Centre for Translational and Molecular Medicine – Translational IT.
DANS	Data Archiving and Networking Services, Dutch KNAW-institute for permanent access to digital research data.
DM	Section Datamanagement, (Research Support), location VUmc.
DMP	Data Management Plan: a document in which the data management process is described.
DOI	Digital Object Identifier: a unique identification number that makes the study findable on the internet, when published in a repository.
DPIA	Data Protection Impact Assessment.
DPO	Data Protection Officer.
EPD	*Elektronisch Patiënten Dossier*, Electronic Health Record (e.g. Epic).
GCP	Good Clinical Practice; legal requirements for conducting clinical trials with a medicinal intervention.
Identifiable / identifier	Any uniquely identifying number, feature or code that can, without considerable effort, potentially identify a specific individual.
ISO14155	Guideline for the clinical investigation of medical devices of the International Organization for Standardization.
Metadata	All the information is necessary to interpret, understand, and use a dataset and to understand raw, intermediate and result data. They concern unambiguous descriptions of the data of the entities in a study.
METC	*Medisch Ethische Toetsings Commissie*, Institutional Review Board, location AMC, location VUmc.
NFU	*Nederlandse Federatie van Universitair Medische Centra*, The Netherlands Federation of University Medical Centres.
Persistent identifier	A long-lasting reference to a study, its documentation and/or its data within a repository, e.g. a DOI.
PIN	Patient Identification Number: a unique, anonymous number for a study participant.
PI	Principal Investigator
PI	Lead of a research line.
Coordinating PI	The investigator who is the final responsible person for the entire study. In case of multicenter studies, the coordinating PI also carries final responsibility for the coordination of the study in the participating sites. *This is a study specific role and **not** related to the function of PI.*
Local PI	Local principal investigator: the main responsible for the study for a site that participates in a multicenter study. *This is a study specific role and **not** related to the function of PI.*
RDNL	Research Data Netherlands; national coalition of data archives that promotes long-term archiving and reuse of research data.
SOP	Standard Operating Procedure: uniformly written procedures with detailed instructions.
Sponsor	*Verrichter;* the person, company or organisation that takes the responsibility for the initiation, organisation and/or financing of a study.
SURF	*ICT-samenwerkingsorganisatie van het onderwijs en Onderzoek in Nederland*, Collaborative organisation for ICT in Dutch education and research.
WGBO	*Wet op de Geneeskundige Behandelingsovereenkomst,* Medical Treatment Agreement Act.
WMO	*Wet Medisch Wetenschappelijk Onderzoek met Mensen*, Medical Research involving Human Subjects Act.
Version control	Applying meaningful version numbers in the filename so that previous and current versions can be distinguished. Use e.g. version 0.1, 0.2, v1.0, v1.1 etc. Or use dates for systematic version naming.

Objective

Research data management aims to control the entire data management process along the research lifecycle; from study preparation, data acquisition, data processing and statistical analysis, writing and publishing, to archiving and open data. The data lifecycle should be well documented, transparent and traceable. Legislation and growing emphasis on issues such as reproducibility and integrity of research require specific working procedures and facilities. Therefore continuous attention for data management is necessary in order to maintain and to improve the current quality standards. For more information see also the Handbook for Adequate Natural Data Stewardship (HANDS) of the NFU.

This Standard Operating Procedure (SOP) is an elaboration of the Amsterdam UMC Research Code, especially chapter 7. The requirements originate from legislation, Amsterdam UMC policies and from good research practice in Amsterdam UMC.

This SOP is accompanied by a template for a Data Management Plan (DMP), location AMC or location VUmc. The DMP allows to document the topics mentioned in this SOP in the context of the study. This SOP refers to several templates, documents and websites that offer standards or additional information.

Scope

This SOP applies to all researchers of the Amsterdam UMC conducting research involving human subjects:

- WMO research, i.e. research that includes subjecting persons to interventions or imposing a particular course of conduct upon them.
- Research with registered or non-registered medicinal products or devices that additionally have to comply with GCP or ISO14155;
- Non-WMO research, i.e. all other research with (groups of) human subjects. This includes qualitative research with unstructured data.

Preclinical research, e.g. on animals, cell lines or body material, is beyond the scope of this SOP.

Responsibilities

This SOP is developed by the Clinical Research Unit (CRU) at location AMC and the section Datamanagement (DM) at location VUmc and is authorised by the Board of Directors. The heads of all Amsterdam UMC departments are responsible for adherence to this SOP and must ensure that their employees, involved in research with human subjects, are informed of this SOP.

Although the Amsterdam UMC as sponsor carries the formal responsibility for all Amsterdam UMC-initiated studies, the final responsible of a study fulfils the role of co-ordinating PI. He or she can delegate tasks, e.g. on data management, on completion of the data management plan, on data processing, analysis etc. to others and should maintain an overview of these roles.

Procedures

The working procedures presented in this SOP are divided into the five phases of the Research Lifecycle: study preparation, acquiring data (or samples), processing & statistical analysis, writing & publishing and archiving & open data. Per phase one can find instructions on how to organize the data management process. Underlined text refers to helpful websites or to Amsterdam UMC documents that can be found in Kwadraet or via the CRU intranet page (location AMC), and in Kwaliteitsnet or via the Datamanagement and/or CRB Intranet pages (location VUmc).

Getting started: Data Management Plan

A Data Management Plan (DMP) must be written using the generic Amsterdam UMC DMP template for location AMC or location VUmc. The DMP must be initiated at study start and completed for each section of this SOP. It can be updated on a continuous basis and it is placed under version control.

- For researchers who work on a ZonMw grant, completion of the Amsterdam UMC template is sufficient. For more information see the ZONMw site 'Toegang tot data'.
- Researchers who work on a Horizon2020 or a *Hartstichting* grant can contact the Research Data Management helpdesk for additional DMP obligations. With regard to special requirements for EU grants contact at location VUmc the Research Grants Desk.

The Research Data Management helpdesks (rdm@amc.nl; datamanagement@vumc.nl) can be contacted for support on the completion of a DMP.

Phase 1: Study preparation

Privacy and security safeguards	
1.1	The sponsor and the coordinating PI must ensure that the processing of the data takes place in accordance with the requirements of the applicable privacy regulation and the privacy policies of Amsterdam UMC.
1.2	A dataset should be set up as little identifiable as possible. However, entirely anonymous data sets will exclude certain forms of data reuse. No directly identifiable data are allowed in the research dataset, unless absolutely necessary and after consulting the DPO AMC or the DPO research, VUmc. *More information on privacy and how to de-identify a data set can be found on the CRU website (location AMC) or on the DM intranetpages (location VUmc).*
1.3	A DPIA should be performed, concerning privacy risks and data protection safeguards with regard to the acquisition, processing and storage of the study data. *At location AMC consult the DPO AMC for details on performing a DPIA. At location VUmc a DPIA is part of the METc intake procedure for both WMO and non-WMO research.*
1.4	Any data collection that can be linked to individual subjects, directly or via a subject identification log, should be registered. At location AMC, see the DPO intranet site. Registration at location VUmc starts here. VUmc researchers need to prove their registration when applying a protocol to the METc for WMO, Biobank and non-WMO studies.
1.5	No research participant can participate in a study without an Informed Consent, unless it is completely anonymous data, or when one of the WGBO exceptions applies. Collecting or reusing the subject's encoded or directly identifiable data must be part of this consent. Information covers the description of the dataset, the time span of data retention, and, if applicable, information on sharing data or on making data available for future follow-up research. Informed consent is either part of the regular WMO Informed Consent procedure, or obtained in a separate procedure.
1.6	At location AMC, all WMO-compliant research should follow the required procedures for approval: • See the CRU website for an overview of SOPs • The METC website or CCMO website. At location VUmc, all WMO-compliant and all non-WMO research, should follow the required procedures for approval. For more information and procedures to follow see the METC website or CCMO website.
1.7	Any file or system, containing identifying data, should be separated from the non-identifying documents (see also 1.32). For location AMC, this is described in *SOP CTR 006 Kw Het verzamelen, filen en archiveren van essentiële documenten* and for location VUmc in the information about the content of the Trial Master File and Investigator Site File.
1.8	In international studies, the coordinating PI should adhere to the regulatory requirements in any of the countries that participate in the study.
Data acquisition	
We distinguish three types of data acquisition: • ***Reuse of existing data*** *that have been set up for another purpose than the applicable study, e.g. third party databases or registries, or the EPD.* • *Using* ***measured data*** *that are generated specifically for the applicable study by a machine or device: think of laboratory assessments, DNA sequencing, (radiologic) images, ECGs, etc.*	

• ***Data collection*** *for which an own data collection tool has to be set up. Examples are a paper or electronic CRF, or an (online) questionnaire.* *Looking up existing EPD data and capturing these in an own data collection tool should be considered both as reuse of existing data and collected data.*	
General	
1.9	The dataset that is acquired must be clearly related and limited to the scope of the study.
1.10	Provide an overview of the data per type of data acquisition and distinguish further into e.g. 'patient characteristics', 'lab data' or 'CT-scans'. Where possible, define the dataset in line with existing standards, both in generic terms or discipline-specific. Terminology standards, classifications (e.g. ICD-10) or data definitions from leading registries or studies (e.g. a minimal dataset) within the applicable are considered and where possible adapted. This will enhance the acceptance and reusability of the dataset.
1.11	If more than one data file or data collection system is merged, the same unique identifier (PIN) is applied in each system in order to link the data in a later stage.
Reuse of existing data	
1.12	Consider whether any existing source of data can be reused. (Part of) the applicable dataset may already be available in existing research databases (registries, public databases) or in the *EPD* (e.g. Epic). *At location AMC, see intranet for example sources of research data.*
1.13	When reusing care data at location AMC, follow the SOP RDM002 Reuse of care data for the purpose of research. *See the CRU website for more details.* For the application procedure for reusing medical data for scientific research at location VUmc see the Data Management intranet pages.
1.14	When reusing any non-Amsterdam UMC data, the conditions for reuse, including the subjects' consent, should be described in a formal agreement with the owner of the data, unless the data can be reused without restriction. *At location AMC, consult the department Legal Research Support for setting up conditions for reuse of data. Specific procedures apply for reusing controlled access (genomic) data.* *At location VUmc, consult Innovation Exchange Amsterdam for setting up conditions for reuse of data.*
1.15	A party that delivers encoded data is responsible for maintaining a subject identification log (see also 1.32). *Use the templates for a subject identification log at location AMC or at location VUmc*
Measured data	
1.16	If the tool for measured data generation will be hosted by Amsterdam UMC or by an Amsterdam UMC-supported external partner, security safeguards can be assumed sufficient. However, if this system is hosted by non-Amsterdam UMC partner or in case a supplier, sponsor, or other research party will have access to the data or receive part of the data, the necessary contractual arrangements regarding careful processing, including technical security safeguards should be made. *At location AMC, consult the department Legal Research Support; at location VUmc, consult Innovation Exchange Amsterdam when it is already clear what kind of agreement is needed. Otherwise ask the DPO research for advice.*
1.17	A description of the measured data (data format and definition) is available.
1.18	All users are trained in working with the system and this has been documented. *At location AMC, use the template CTR F13 Study Training Log; at location VUmc use the Trainingslog template.*
Data collection	
1.19	When making a paper CRF (pCRF), use standard pages (e.g. for medical history or adverse events) where possible. It is however strongly advised not to use paper CRF's anymore but to work with an eCRF. *At location AMC, see the CRU website for an example CRF, standard CRFs and a checklist for pCRF design.* *At location VUmc see the SOP CRF's ontwerpen, invullen en invoeren.*
1.20	It is strongly recommended to use software that is validated and designed for data collection. • Preferably use software system, supported by Amsterdam UMC or Amsterdam UMC-partners (e.g. Castor, OpenClinica, LimeSurvey or Survalyzer). • For WMO studies, the use of GCP compliant data management software like Castor is mandatory. • The use of spreadsheets, such as Microsoft Excel, is not allowed. • The use of statistical packages for data collection (e.g. SPSS, R or STATA) is only allowed in non-WMO

	studies with limited data collections (indication: < 25 subjects and < 25 variables). *See the CRU website (location AMC) or the Data Management website (location VUmc) for an overview of applications for randomization, eCRF or web questionnaire design, for dataflow control and for the support you can obtain.* *Consult the Research Data Management helpdesks (rdm@amc.nl; datamanagement@vumc.nl) to check whether a company or organization is an Amsterdam UMC partner.*
1.21	If the data collection system is hosted by Amsterdam UMC or an Amsterdam UMC-partner (e.g. Castor, OpenClinica, LimeSurvey, Survalyzer, CTMM-TraIT, SURF, Durrer Center, NFU Teamwerk collaboration platform), security safeguards have been verified. However, if the data collection system will be hosted by a non-Amsterdam UMC-partner or in case a supplier, sponsor, or other research party will have access to the data or receive part of the data, the necessary contractual arrangements regarding careful processing, including technical security safeguards should be made. *At location AMC, consult the department Legal Research Support; at location VUmc, consult Innovation Exchange Amsterdam.* *Consult the Research Data Management helpdesks (rdm@amc.nl; datamanagement@vumc.nl) to check whether a company or organization is an Amsterdam UMC partner.*
1.22	Describe the data definitions in a data dictionary or codebook. *Use meaningful variable names that can easily be understood by future users of the dataset.* *Examples templates for location AMC and for location VUmc can be used.* *When using Castor for data collection, an overview of all fields including data format can be automatically generated. When working with OpenClinica for data collection, the Excel files that define the CRFs can serve as a data dictionary.*
1.23	Procedures to check for completeness, correctness and consistency of the data are in place and are documented. *Check individual data fields, using **univariate** validation checks that aim at checking, such as the unicity of key fields (e.g. PIN), upper and lower bounds and type of missing data ('not done', 'unknown', 'completion pending'). Instead of free text fields, define code lists to enforce that only predefined answers are chosen. Document these univariate checks in the data dictionary templates for location AMC and for location VUmc.* *Check consistency between multiple fields, using **multivariate** validation checks such as the start date of a visit or Adverse Event that cannot be after its stop date. These consistency checks must be documented, e.g. in the data validation plan templates for location AMC and for location VUmc.*
1.24	The system should be tested before it is being used and by someone not involved in the development of the system. Type of tests, test results, follow-up of the findings and final approval must be documented. *For documentation of system testing, templates are available for location AMC and for location VUmc.*
1.25	Access to the data collection system is only granted after approval of the coordinating PI and only to personnel that is involved in the study. It is based on unique user identification, since shared login names cannot be traced back unambiguously to a person. When no longer required, access is revoked. Keep documentation of granting, modifying and revoking study access (date, role within the system, read or write access, approval of the coordinating PI). *For this purpose, a template is available for location AMC and for location VUmc.*
1.26	The data collection system should log the identity of the persons using the system.
1.27	All users are properly trained in using the system. This is documented in the training log templates for location AMC and for location VUmc.
1.28	For each study-specific data collection, source documentation is available, unless the absence of source documentation has been described in the study protocol or the Data Management Plan.
Data storage	
Besides the tools that are used to acquire data, storage of the generated data during the study is also important.	
1.29	All data storage should take place on one of the following back upped drives: • The department's G-drive (location AMC) or M-drive (location VUmc). Create an own study folder here. Data that are not shared with colleagues can be placed on the personal network folder (H-drive (location AMC) or N-drive (location VUmc). • For large amounts of data and the archiving of raw data and research output, the Store4Ever storage system is available. *For creating a study folder on the G-, H-, M- or N-drive, or for access to Store4Ever, contact the ICT*

	helpdesks of location AMC or location VUmc. • Provide a well-organized folder and file structure with the correct access rights and document this. For more information at location VUmc see the SOP *Studiespecifieke mappenstructuur.* For WMO research, be aware that it is not allowed to keep only a digital version of the Trial Master File. • On a (database) server that falls under the central backup regime. • On a (database) server of an Amsterdam UMC-partner (e.g. Castor, OpenClinica, Survalyzer, SURF, CTMM-TraIT, Durrer Center, NFU Teamwerk collaboration platform) • Outside the Amsterdam UMC at a NEN7510 and/or ISO27001 certified data centre: o In absence of such certification, one should be able to demonstrate that the processing of the data takes place in accordance with the security requirements of the applicable law. *Contact the DPO AMC or the DPO research, VUmc for further consultation.* o The necessary contractual privacy arrangements regarding careful processing, including technical security safeguards should be made. *At location AMC, contact the department Legal Research Support; at location VUmc, consult Innovation Exchange Amsterdam.* This implies that neither the C-drive of one's own computer, nor any stand-alone device such as an external hard drive, personal laptop or USB stick is used. *Consult the Research Data Management helpdesks (rdm@amc.nl; datamanagement@vumc.nl) to check whether a company or organization is an AMC or VUmc partner.*
1.30	Sufficient capacity and financial means are available for data storage, both during the study as well as for long term archiving (see Phase 5). *For estimating the file size of the dataset, either look at the size of a comparable file, or make an estimation of the data size at subject-level and multiply this by the number of subjects. At location AMC, see the ADICT website for costs of storage. At location VUmc, the use of the M- and N-drive is free of charge. For costs of Store4Ever storage see the VUmc ICT intranet site.*
1.31	Only authorized staff has access to the data and when no longer required or permitted, access is revoked. The coordinating PI should keep track of authorizations and documentation. *Contact the ICT helpdesks of location AMC or location VUmc for granting or revoking authorizations on Amsterdam UMC-owned storage locations.*
Subject identification log	
1.32	The subject identification log should be kept separate from the patient data as described for location AMC in the *SOP CTR 006 Kw Het verzamelen, filen en archiveren van essentiële documenten* and for location VUmc in the information about the content of the Trial Master File and Investigator Site File.
1.33	In multicentre studies, a site-specific subject identification log should be kept on each study site. It is not allowed to store an overall multicentre log at the coordinating center, unless consent has been given by the research participants. *Use the templates for a subject identification log at location AMC or at location VUmc.*
Data sharing	
1.34	In case the project is carried out in a consortium with two or more partners, written agreements on data management, privacy and intellectual properties have to be made. *For setting up such agreements, contact the department Legal Research Support at location AMC; at location VUmc, contact Innovation Exchange Amsterdam.* • Anonymous data can be shared according to these agreements. • Identifiable and encoded data can only be shared with the study subject's written consent.

Phase 2: Data collection	
General	
2.1	Record the personnel involved in data collection on the templates at location AMC and at location VUmc.
2.2	When transcribing from pCRF into an eCRF, the original data sources should be retained in the paper study file.
2.3	In the case of a double blind study: when deblinding a single subject, permission of local PI is required, as well as detailed documentation. *Use the template RDM F05 Deblinding randomization code for deblinding at location AMC or at location*

	VUmc.
Including a new subject	
2.4	It is only allowed to acquire data from patients that gave informed consent.
2.5	Data should only be traceable to the study subject through a unique PIN. *For more information on de-identification of study data, see at location AMC the CRU website and at location VUmc the datamanagement site.*
Reuse of existing data / use of measured data	
2.6	Save the dataset that has been acquired as a read-only file. Use a copy of this file for further processing and statistical analysis.
2.7	Validate the dataset on completeness, correctness and consistency before processing the data (see also item 1.23).
Externally acquired data	
2.8	When an external party delivers the dataset and it contains identifiable data, this party should remove or encode it before delivery.
2.9	When the dataset needs to contain identifiable data, contractual arrangements should be in place. These arrangements describe the responsibility of the external party to ensure that the collection ofthe dataset and making the dataset available, including the actual delivery, is in accordance with the data privacy regulation (e.g. on informed consent).
2.10	Data are transferred in a secure way or via SURFfilesender or using the *'Veilig verzenden'* button in Outlook.
Data entry	
2.11	The local PI is responsible for data entry or modification in the pCRF or eCRF. In the case of multicentre studies, the coordinating PI may only enter data after his or her authorization. This must be documented on the delegation log templates at location AMC and at location VUmc.
Quality control	
2.12	Document the study procedures so that data collection is done in a standardized and transparent way. This must be documented in the protocol and, at location VUmc, in the *Study operations manual*.
2.13	Apply all available procedures to check the dataset on completeness, correctness and consistency as defined during study preparation (see also item 1.23).
2.14	Completion of the data collection is signed by the coordinating PI, either electronically or manually. In the case of multicentre studies, the site specific data collection is signed by the local PI. *This does not apply to patient reported outcomes, such as questionnaires.* *eCRF systems such as Castor and OpenClinica provide user roles for approving data entry.* *Use the template CTR F10 Principal Investigator's Statement for manual signing at location AMC or at location VUmc.*
Change control	
2.15	After first entry, changes and deletions in the **data** have to be tracked in an audit trail: • Changes in data on a paper document (e.g. pCRF) should be dated, initialled, and explained (if necessary) and should not obscure the original entry. • Data collection systems that contain an audit trail, perform this tracking automatically. *eCRF systems such as Castor and OpenClinica contain an audit trail on data changes.* • In all other cases the changes are actively documented. *Use the template RDM F04 Change management to document changes in the data at location AMC or at location VUmc.*
2.16	After starting the data collection, document any changes in the **design** of the system: • Changes in the pCRF are clearly described and result in a new version. • Data collection systems that contain an audit trail, automatically perform version control. *eCRF systems such as Castor and OpenClinica and contain an audit trail on eCRF changes.* • In all cases, use the templates at location AMC or at location VUmc to document changes in the design of the data collection system, even when this tool already keeps an audit trail.

Phase 3: Data processing & statistical analysis	
Locking a data collection system *This section can be skipped in case of interim (pre-final) analyses.*	
3.1	When all data are completed and corrected, the data collection system should be locked before starting processing and analysing the data. *If no locking functionality is available, lock the system by setting it read only, revoking write-access, or removing it from the original, accessible location. At location VUmc, see the SOP Sluiten en Ontsluiten database.*
3.2	The coordinating PI must give approval for locking an system. *Use the template RDM F07 Lock study data for documentation at location AMC or at location VUmc.*
Export to the data processing and statistical environment	
3.3	Export the data preferably to a general accepted system for processing and statistical analysis, such as SPSS, R, SAS or STATA. If this is not possible, the data should also be stored in a generic format that is machine actionable, irrespective of the software provider. Follow the DANS guidelines for referred and acceptable data formats.
3.4	Record the software tools used for data processing and statistical analysis, including version number, in the DMP or in the Statistical Analysis Plan.
Data processing and statistical analysis	
3.5	Do not use the original exported dataset for processing and statistical analysis, but store it as a read only file with a new name and in a separate folder. For further processing and statistical analysis, make versioned working copies.
3.6	It should always be possible to compare the processed data with the original data. Therefore, all data transformations, restructuring and statistical analyses should be programmed and filed in syntax or script.
3.7	Add descriptive comments to the syntax or script that explain what it achieves and give the file a meaningful file name. It should be obvious which syntax or script corresponds with which part of the data processing.
3.8	Place all processed datasets and syntaxes or scripts under version control. Future users should be able to reproduce the (published) study results.
3.9	If any corrections on the original data are needed in this phase, make these in the original source (e.g. the Castor eCRF) and log them as demanded in section 2.15. If this is not possible, program these corrections within the processing or statistical environment, using a syntax or script file that is saved and can serve as a change log.
3.10	In case of export for pre-final analyses, randomization results can be fictitiously assigned.
Sharing for processing or statistical analysis	
3.11	Only share data from subjects that gave their informed consent with external parties, (for example with an external statistician).
3.12	In the case of multicentre studies, the local subject identification log must remain on site and under control of the local PI.
3.13	Data are transferred in a secure way or via SURFfilesender or using the *'Veilig verzenden'* button in Outlook. Not-confidential data and documents can also be shared via Surfdrive.

Phase 4: Writing & publishing	
Filing	
Although in this phase the manuscript is written and published, journals require that underlying data and analysis scripts that lead to the reported results are available. Therefore, organise the datasets and files, used for publication, in such a way that the presented results can always be reproduced.	
4.1	All datasets, subject selections, methods and syntaxes or scripts **at the moment of submission** of a publication must be clearly linked to that publication. For each publication a folder is created that contains: • the final pre-processed data • a list of screened and excluded subjects, containing PIN and reason for exclusion • annotated syntax and scripts files, including name and location of the relevant data file. • the relevant processed data files • the statistical output as reported in the manuscript

	• the final, submitted manuscript and revised manuscripts.
Accessibility of the dataset	
4.2	Pre-register the study in a catalogue or repository, create a Persistent Identifier (DOI, Handle) and include this the publication. See Phase 5 for more details.

Phase 5: Archiving & open data	
Open data	
After publication of the manuscript, the study data can be made accessible for reuse by other researchers. We recommend to do this, unless there are intellectual or commercial property issues that forbid this. Conditions for reuse and/or an embargo period can be defined. For publishing the data, one can make use of catalogues or repositories, that allow to present the study, including (optional) documentation and data on a website. In the next section the relevant procedures are provided. *Qualified data repositories automatically create a so called Persistent Identifier, e.g. a Digital Object Identifier (DOI) that refers to the registration in the repository. By including this Persistent Identifier (e.g. DOI) in a publication, a link between publication and dataset is created.*	
5.1	Publication of data, unless they are entirely anonymous, must be covered by the patient's informed consent and may not conflict with legal, contractual or other provisions.
5.2	Set up a document with conditions that have to be met before the data can be reused by others. Think e.g. of data security, statistical analysis, publication and authorship, sharing or linking data legal or financial issues. *For templates and advice on conditions for reuse, consult the department Legal Research Support at location AMC; at location VUmc, contact Innovation Exchange Amsterdam.*
5.3	Consider whether an embargo needs to be observed, during which the data will not be accessible for reasons of publication, public safety, privacy, intellectual property or commercial interests. The embargo period is maximum three months (nine months for patents). Explain the reasons for the embargo and its period.
5.4	Consider which data (raw or pre-processed data, which stages of processed data) will be published, in the context of file size and anonymity of the dataset. Think of large genomic data files or of interview video tapes which are identifiable by definition. Include the location of the published data. *To help you determine which research data constitute valuable source material for further research, both DANS and RDNL offers tools for the selection of research data for long term archiving.* *Consult the Research Data Management helpdesks (rdm@amc.nl; datamanagement@vumc.nl) for advice.*
5.5	Publish the study and/or data in a catalogue or repository, preferably certified (with a CoreTrustSeale) that automatically creates a Persistent Identifier (PI), e.g. a DOI. Decide whether the data are published or only a contact address for requests for data reuse is provided. To ensure long-term usability, accessibility and preservation of data, you have to make use of a 'preferred' file format, preferably machine readable. DANS has made an extensive list of preferred formats. *Consult the Research Data Management helpdesks (rdm@amc.nl; datamanagement@vumc.nl) for advice on publication of the study, its documentation and data.* *Consider the BBMRI catalogue for sharing biobank data or www.re3data.org for field specific repositories.*
5.6	Provide descriptive documentation (metadata) of the **study**. Decide what is necessary for researchers to understand the study. Choices will depend on the research topic; think of a study description according to general accepted standards, at least consisting of the study protocol, Statistical Analysis Plan, and DMP. *So called metadata schemes offer high level study descriptions in a generic and standardized way. Consult the Research Data Management helpdesks (rdm@amc.nl; datamanagement@vumc.nl) for advice on metadata schemes.*
5.7	Provide descriptive documentation (metadata)of the **data**. Decide what documentation is necessary for researchers to reproduce and reuse the data. Choices will depend on the research topic and type of datasets, think of a data dictionary, data validation plan, coding and syntaxes or scripts used for analysis, etc. *The way data are documented will differ per discipline. Consult the Research Data Management helpdesks (rdm@amc.nl; datamanagement@vumc.nl) for advice on these metadata. For more information on metadata standards see also FAIRsharing.org.*
Transfer to an external party	
5.8	In the case the study data and documentation will be transferred to an external party, a Data Transfer Agreement should be set up.

	For setting up such an agreement, consult the department Legal Research Support at location AMC; at location VUmc, contact Innovation Exchange Amsterdam.
5.9	Data or documentation should not leave Amsterdam UMC; in case of transfer to an external third party, always keep copies of the transferred files, including documentation of the transfer.
5.10	Data and documents are transferred in a secure way or via SURFfilesender or using the *'Veilig verzenden'* button in Outlook.
Digital archiving	
5.11	For WMO-compliant studies, follow the procedures for what to store in the archive at location AMC and at location VUmc. If the sponsor of the study is an academic hospital, the retention period is 15 years for WMO and non-WMO studies. However, the researcher may agree a different retention period with other types of sponsor. *Also archive essential e-mails. In Microsoft Outlook, select the relevant e-mails, click the right mouse and convert to pdf.*
5.12	Allocate sufficient budget for archiving the data (see also topic 1.30).
5.13	Document the location of the digital archive in the DMP. *At location AMC, use the L-drive for long-term storage and for location VUmc use the Store4Ever storage system. Contact the ADICT-helpdesk for creating a folder on this L-drive and for access to store4Ever at location VUmc contact the ICT Servicedesk*.
5.14	Keep the subject identification log only if this is necessary for future reuse or data linkage and does not conflict with the subject's informed consent. This log should be kept separate from other patient data as described at location AMC in the *SOP CTR 006 Kw Het verzamelen, filen en archiveren van essentiele documenten* and at location VUmc in the information about the content of the Trial Master File and Investigator Site File. In multicentre studies, the site-specific subject identification logs should be archived on each study site. It is not allowed to centrally archive an overall multicentre log across the sites.
Paper archiving	
5.15	Document the physical location of the paper archive and the retention period in the DMP.
5.16	Follow the guidelines for archiving paper documents of WMO-compliant studies at location AMC and at location VUmc.