

Research Data Management & Open Data

Rudy Scholte, Joost Daams - June 10, 2021





Research data management

Concerns the organization of data...

... from its collection, through the **research lifecycle** to the dissemination and archiving of valuable results.





Goal of RDM

Deliver reliable and reusable data

documentation

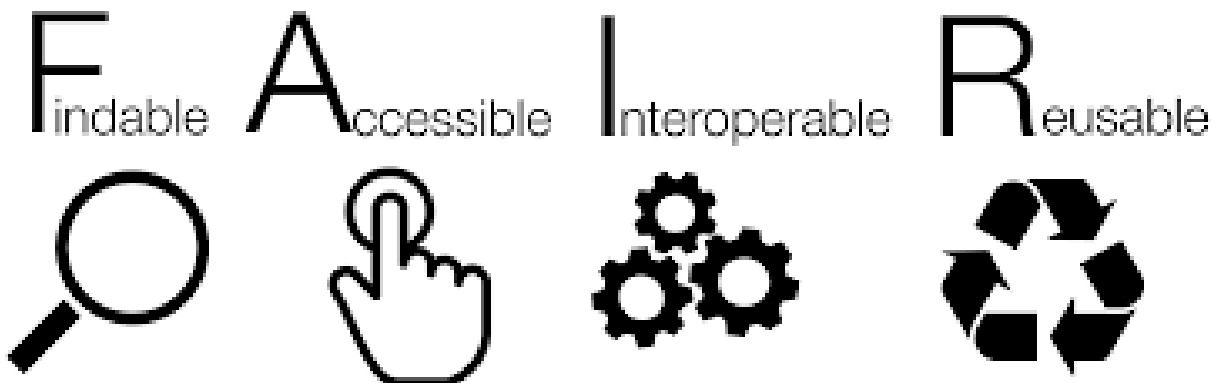


transparency



reproducibility





- Findable** make metadata of your study (online) available
- Accessible** make your data available for others (under conditions)
- Interoperable** use open standards and definitions that allow reuse by others
- Reusable** provide metadata so that others can reuse your data

See Wilkinson et al, The FAIR Guiding Principles for scientific data management and stewardship, <http://www.nature.com/articles/sdata201618>



Increasing demands for RDM



Avoid data leaks: de-identification, storage facilities, consent procedures



Data management plan, open data



Reproducibility





Amsterdam UMC working procedure for RDM

Standard Working Procedure (SOP)
Data Management Plan

Scope: research with human subjects
(to be extended to all research)

See

- [AMC CRU intranet](#)
- [VUmc Datamanagement intranet](#)
- K2

The image shows the cover page of a document titled "Research Data Management Standard operating procedure". At the top left is the Amsterdam UMC logo with the text "Amsterdam UMC University Medical Centers". The title "Research Data Management" is in a large, bold font, followed by "Standard operating procedure" in a smaller font. Below this is a table with two columns: a label column and a value column. The table contains information about the SOP-id, version, date effective, authors, advisory board, authorized by, signature, date, and released by.

SOP-id	RDM001
version	4.0
date effective	January 1, 2020
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	Expertise group 'Data Matters'
Authorized by	J.A. Romijn, CEO, location AMC C.H. Polman, CEO, location VUmc
signature	
date	
Released by	M.J.A.P. Daemen, Chair Amsterdam Research Board



Getting started: Data management plan

- Documentation of your data management process
- Required by SOP, funders
- DMP replaces ZonMW DMP

Support via rdm@amsterdamumc.nl

Reuse of existing data <input type="checkbox"/> Not applicable	
1.12	Specify the source that is used to acquire data: ...
1.13	<input type="checkbox"/> The reuse of data is covered by the subject's informed consent
	<input type="checkbox"/> Contractual arrangements are in place for reuse of data from an external party; <i>provide document name and location: ...</i>
	Additional information:
Measured data <input type="checkbox"/> Not applicable	
	<input type="checkbox"/> The system that generates the measured data are hosted by the Amsterdam UMC or by an Amsterdam UMC partner, <i>namely: ...</i>
	<input type="checkbox"/> The system that generates the measured data is hosted by a non-Amsterdam UMC partner, <i>namely: ...</i>
	<input type="checkbox"/> Contractual arrangements are in place for use of measured data from an external party
	<input type="checkbox"/> A description of the generated data is available; <i>provide document name and location: ...</i>
	<input type="checkbox"/> All users are trained in the system and this has been documented; <i>provide document name and location: ...</i>
	Additional information:
Data collection <input type="checkbox"/> Not applicable	
1.20	<input type="checkbox"/> The data collection system is hosted by the Amsterdam UMC or by an Amsterdam UMC partner, <i>namely: ...</i>
1.21	<input type="checkbox"/> The data collection system is hosted by a non-Amsterdam UMC partner, <i>namely: ...</i>
	<input type="checkbox"/> Contractual arrangements are in place for data collection through an external party
1.22	<input type="checkbox"/> A data definition (data dictionary) is available; <i>provide document name and location: ...</i>
1.23	<input type="checkbox"/> Checks on completeness, correctness and consistency are incorporated in the data collection system and have been documented; <i>provide document name and location: ...</i>
1.24	<input type="checkbox"/> The system for data collection has been tested: <i>Last test date: (DDMMMYYYY) By: (name and function) Version tested:</i>
	<input type="checkbox"/> The test findings and final approval are documented; <i>provide document location: ...</i>
1.25	<input type="checkbox"/> Access to the data collection system is managed by the coordinating PI. User roles and authorizations are documented; <i>provide document name and location: ...</i>
1.27	<input type="checkbox"/> Users are trained in the data collection system; this is documented; <i>provide document name and location: ...</i>
	Additional information:



Privacy & security

Aim at minimal identifiability

- a. anonymous
- b. encoded (pseudonymous, indirectly identifiable)
- c. identifiable

b and c: consent required (\neq WMO consent)

- description data set
- time span of data retention
- sharing data
- reuse in future research

Data Protection Impact Assessment (short version), Registration of your data collection

see intranet pages Data Protection Officer (Functionaris Gegevensbescherming)





Patient recruitment

Automated cohort identification

- Alternative for manual look-up
- SlicerDicer (Epic)
- Research Data Platform (RDP)
- Unstructured data: text searches

Contact rdm@amsterdamumc.nl for request



Data acquisition: reuse of existing data

Research Data Platform (RDP)

For both locations

Set of Detailed Clinical Models
Diagnosis, Medication, Lab

Epic, pre-Epic, non-Epic
e.g. MUSE, Biobank

SOP see intranet CRU-DM, Datamanagement, K2

Requests to rdm@amsterdamumc.nl



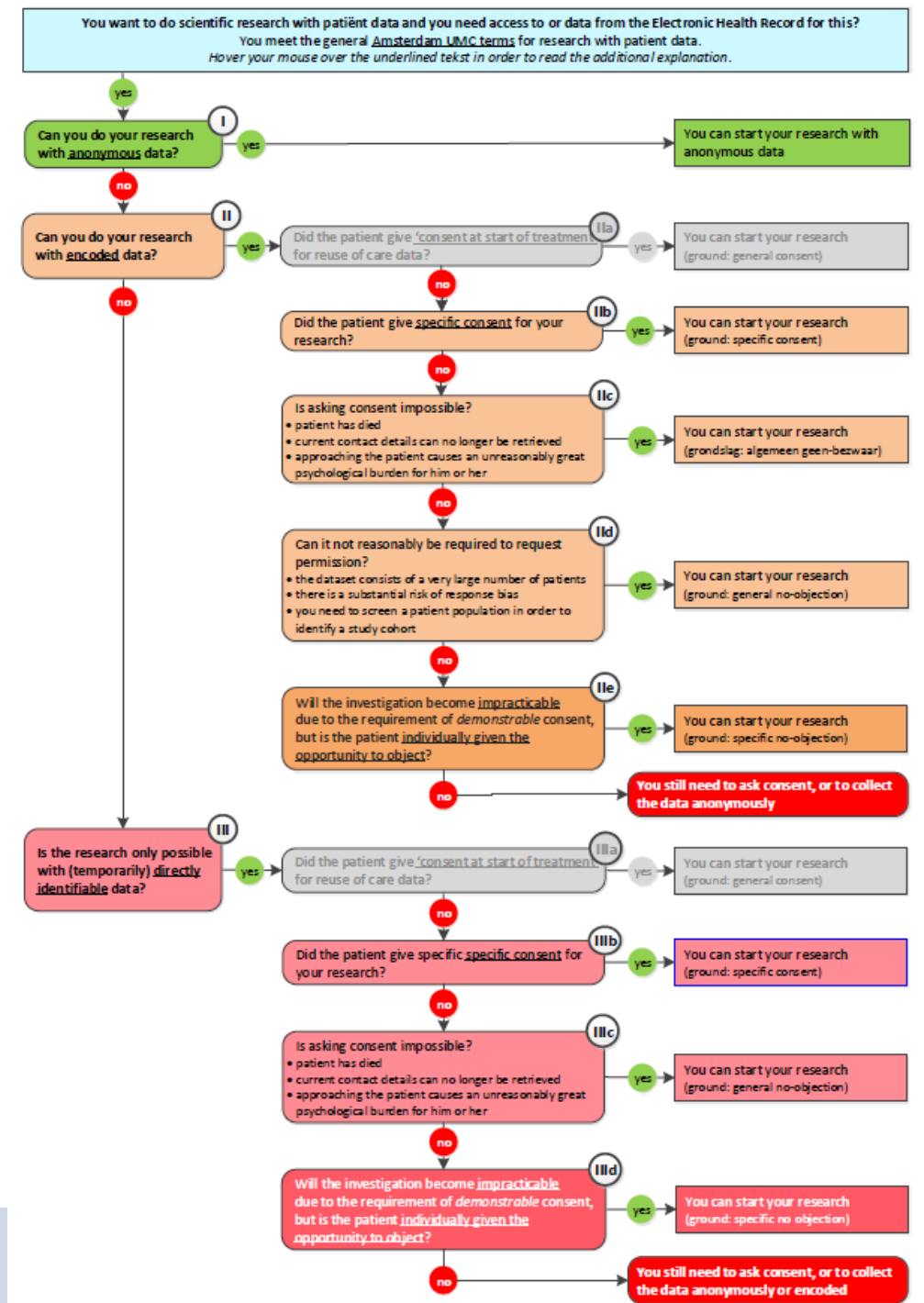
Reuse of care data for the purpose of research

Standard operating procedure

SOP-ID	RDM002
Date effective	October 1, 2020
Author	R.A. Scholte, Head Data Management, Clinical Research Unit, location AMC M.C. Ploem, health lawyer, Amsterdam UMC, location AMC E.J. Kranendonk, health lawyer, Amsterdam UMC, location VUmc
Consulted	Good Research Practice Committee Amsterdam UMC Amsterdam UMC Research Board
Reviewed by	M. Inge, data protection officer Amsterdam UMC, location AMC M. Paardekooper, data protection officer Amsterdam UMC, location VUmc
Authorized by	C.H. Polman, CEO



Amsterdam UMC / AVG-WGBO procedures for reuse of care data





Tools for data collection

Use central supported data management software

Castor (free use), LimeSurvey/Survalyzer, REDCap

otherwise additional requirements, see [intranet page DPO](#)

No Excel for data collection!

Statistical package: only in non-WMO studies with limited data sets

Use 'Amsterdam UMC-approved' tools:





Support on manual data collection tools

Randomization	Castor	-
	ALEA	build
eCRF	Castor, RedCap	test, build
Web questionnaire	Castor	-
	LimeSurvey/Survalyzer	test, build

Information on intranet

AMC: CRU-DM

VUmc: Datamanagement



Tools for manual data collection

Data quality checking

Apply wherever possible:

- *Univariate checks*
- *Mandatory fields (incl coding for missing)*
- *Multivariate checks*
- *Conditional display*
- *Manual (visual) checks*

Document in Data Dictionary / Data validation plan

Testing (colleague, RDM)

Access control





Data storage

At safe location

Amsterdam UMC network

Amsterdam UMC partner

Certified third party + processor agreement

Large volumes of data: consult rdm@amsterdamumc.nl

External access via CDW / VIEW

DMP: Estimate size, reserve €'s





Phase 2

- Under responsibility local PI
- Run your validation checks
- Change control
 - On paper
 - Audit trail
 - Self documented
- Sign off by local PI





Phase 3

- Lock
- Export
- Processing and analysis





Lock data collection

- PI approval



- Documented



- Locking functionality
 - Block access
 - Read only



Export from data acquisition environm

Reusable



Sustainable

DANS preferred formats

Type data	Preferred format(s)	Acceptable format(s)
Tekst documenten	<ul style="list-style-type: none"> PDF/A (.pdf) 	<ul style="list-style-type: none"> OpenDocument Text (.odt) MS Word (.doc, .docx) Rich Text File (.rtf) PDF (.pdf)
Platte tekst	<ul style="list-style-type: none"> Unicode TXT (.txt, ...) 	<ul style="list-style-type: none"> Non-Unicode TXT (.txt, ...)
Spreadsheets	<ul style="list-style-type: none"> PDF/A (.pdf) Comma Separated Values (.csv) 	<ul style="list-style-type: none"> OpenDocument Spreadsheet (.ods) MS Excel (.xls, .xlsx)
Databases	<ul style="list-style-type: none"> ANSI SQL (.sql, ...) Comma Separated Values (.csv) 	<ul style="list-style-type: none"> MS Access (.mdb, .accdb) dBase III or IV (.dbf)
Statistische data	<ul style="list-style-type: none"> SPSS Portable (.por) SAS transport (.sas) STATA (.dta) 	<ul style="list-style-type: none"> R (*)
Afbeeldingen	<ul style="list-style-type: none"> JPEG (.jpg, .jpeg) 	



Data processing & analysis



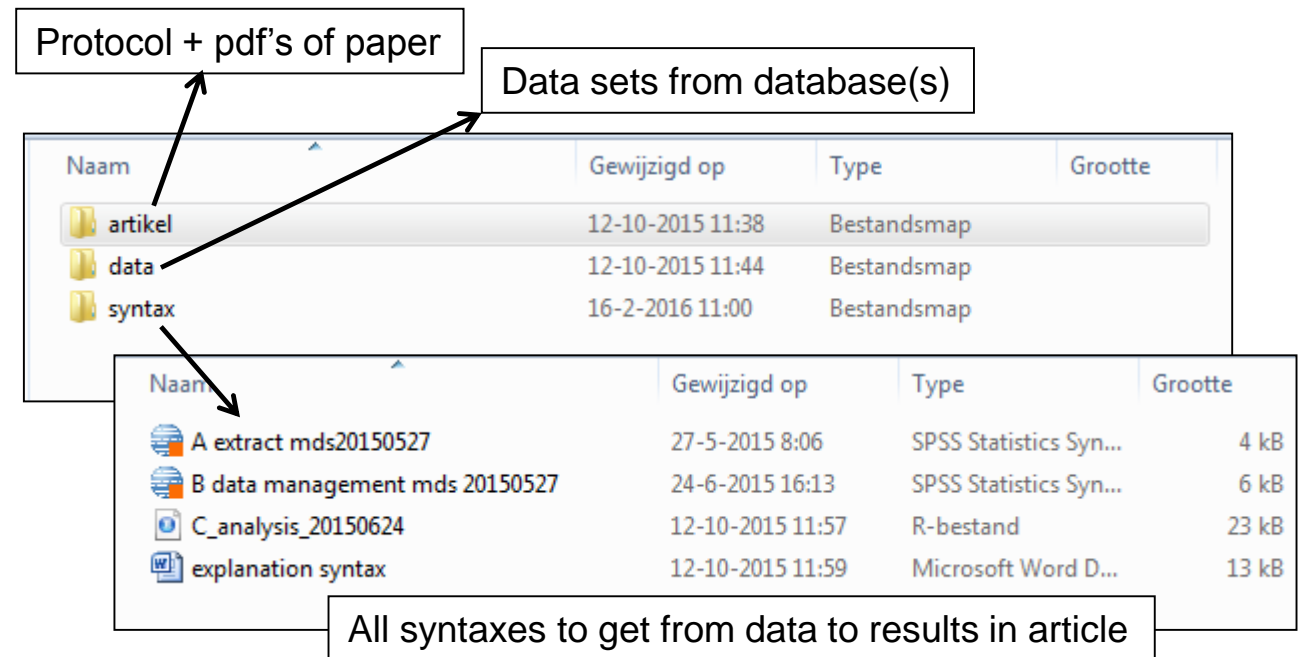
- Keep read-only copy
- Version control
- Syntax files
- Descriptive comments



Writing & Publishing

Phase 4



- Relevant data files
- Scripts and syntax files
- Dropped subjects
- Processed data files
- Statistical output
- Manuscript

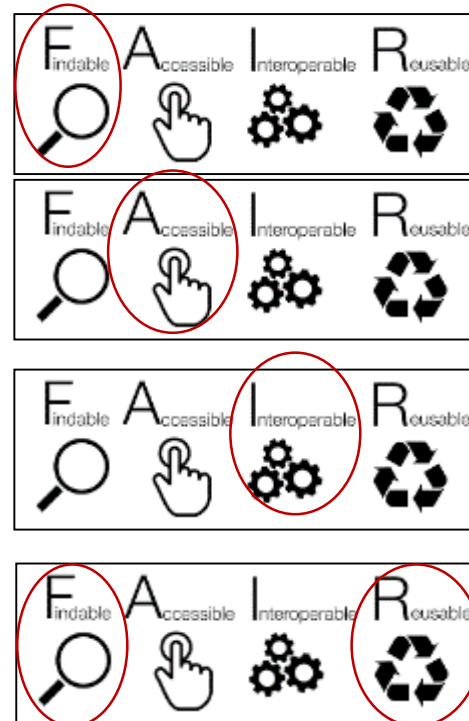




Phase 5

Open data

- Repository  
- Conditions for reuse
- Standard definition, sustainable format
- Metadata at study level, at data level





Archiving

- What data should I archive?

Special attention for

- identifiable data
- subject identification log

- Central archive available via RDM

- provide metadata
- file management via iRODS
- linked to data repository

- SURF facilities, Store4Ever, L-drive



Thank you!

RDM helpdesk: rdm@amsterdamumc.nl

