Plan Overview

A Data Management Plan created using DMPonline

Title: Care2report: The development and testing of automated medical visit summarization as integrated solution to enhance time with patients and reduce administrative burden in clinical geriatrics

Creator: Lourens Kraft van Ermel

Principal Investigator: Sandra van Dulmen

Data Manager: Mercedes Beltrán, Carlijn Hohuizen

Affiliation: Radboud University Medical Center (Radboudumc)

Funder: ZonMw (Nederlands)

Template: Data management ZonMw-template 2019

Project abstract:

Physicians have to make a report of every medical visit. This is very time-consuming. Automated summarization of medical visits could decrease the administrative burden of clinical documentation and leaves more time to spend on patients. In this study, we will investigate whether automated summarization of long geriatric outpatient visits (1.5-2 hrs) is possible and reliable. In addition, we will investigate whether it enhances visit outcomes and if it is useful in other healthcare settings.

In WP1, we will build the software system Care2report up to a minimal viable product guided by the standard geriatric visit protocol, i.e. a Comprehensive Geriatric Assessment. Then, in WP2, the Care2report prototype is tested on effectiveness on summarization time, visit length, objective

report quality, administrative burden, communication process and quality, and patient and healthcare provider (HCP) outcomes. Finally, in WP3, we explore what ethical and practical conditions should be fulfilled to allow other HCPs to benefit from Care2report.

ID: 141796

Start date: 01-05-2023

End date: 30-04-2028

Last modified: 11-01-2024

Grant number / URL: 09120012110006

the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Care2report: The development and testing of automated medical visit summarization as integrated solution to enhance time with patients and reduce administrative burden in clinical geriatrics

1. General features of the project and data collection

1.1 Project leader contact details

Prof. Dr. Sandra Van Dulmen
Coordinator research program Communication in Healthcare at Nivel; endowed professor 'Communication in healthcare', Radboud university medical center, the Netherlands; adjunct professor at the University of Borås, Sweden
email: s.vandulmen@nivel.nl; tel.: 0302729703

1.2 I have composed my DMP with the assistance of a data stewardship (or management) expert. List his or her name, function, organisation/department, phone number and email address.

- The data expert has been trained (and has certification) in data stewardship (please explain).

drs. Carlijn Hofhuizen
Data steward/coordinator ‘ondersteuningscluster panel- en surveyonderzoek’ Nivel
+31(0)302729781
c.hofhuizen@nivel.nl
The data steward participated in the FAIRdata workshops in the ZonMw Covid-19 bottom-up. funds. She was involved as data-steward in two other ZonMw projects. Furthermore, the data steward has 11 years of experience in collecting research data, data quality management, process control and privacy policy.

Mercedes Beltrán, M.I. (Mercedes)
Data steward, Faculty of Science, Utrecht University
+31 6 81031356
m.i.betran@uu.nl
Trained by Research Data Netherlands (RDNL) and Landelijk Coordinatiepunt Research Data Management (LCRDM). GCP Certified (Good Clinical Practice).

1.3 In collecting data for my project, I will do the following:

- Use existing data (please specify)
- Generate new data

This project will re-use existing data (literature review in WP2). Other existing data refer to unpublished pilot data (pseudonymized transcripts of patient interviews) collected by the Informatics department at the UU (Care2report research program) and safely stored at Nivel.

New data will be generated in WP1, 2 and 3. In WP1, data will be collected for building the software system Care2report, in WP2 for performing the RCT study, and in WP3 for examining the interest in Care2report among other HCPs. In WP2, output generated by Care2report will also be compared with conventional reports extracted from the EHRs (Epic) in geriatrics. In the table below, an overview of data to be collected/generated in each WP can be found.

	Data description	Data type	Format	Conversion to open format
WP1	Audio recordings of geriatric outpatient visits	Audio	.wav	.mxf // .mka
WP1	Audio transcripts of geriatric outpatient visits	Text	.doc // .txt	.pdf/a // .txt
WP1	Audio transcripts (pseudonomyzed/masked)	Text	.doc // .txt	.pdf/a // .txt
WP1	Written reports of geriatric outpatient visits * * obtained from electronic patient file	Tabular	.xls	.csv
WP1	Analysis scripts	Scripts	.py	-
WP1	MG ontology	Modelling	RDF/XML	-
WP1	Software documentation	Text	.txt	.txt
WP2	Stata files (questionnaire outcome)	Statistical analysis	.dta	.dat / .DO
WP2	Questionnaires HCP/Patient/Caretaker	Tabular	.doc	.txt
WP2	Video recording; geriatric outpatient visit	Video	.mp4 // .avi	.mxf // .mkv
WP4	Interview managers and HCPs	Text	.doc	.txt

1.4 In my research, I will use:

- A combination of quantitative and qualitative data

1.5 I will be reusing or combining existing data, and I have the owner's permission for that.

- Yes, I have permission to use the data

Data (pseudonomyzed transcripts of recorded medical visits) were obtained using consent as legal basis with permission for data re-use. This is unpublished data collected by the Department of informatics at Utrecht University. The original data are safely stored at Nivel.

There are no technical restrictions nor need for data conversion, data are already in the format needed for re-use.

1.6 In collecting new data, I will be collaborating with other parties.

- Yes, we have reached agreements on the user rights of the data used in the project
- Yes, I will collect the new data in conjunction with other researchers or research groups

Data from geriatric visits, patient and HCP questionnaires, and HCP/manager interviews will be collected at Radboudumc and stored at Nivel. Original video- and audio-recordings will only be accessible from a secure PC on the premises of Nivel.

A technical programmer from Utrecht University will develop and implement a stand-alone transcription and pseudonymization service at Nivel. This transcription and pseudonymization application will be executed for the audio/video data by a Nivel project member at the Nivel building.

For WP1, Nivel will share the pseudonymized transcription of the audio-files with researchers at Utrecht University. After the Care2report prototype is tested at the UU a final report will be shared with Nivel and other consortium partners.

For WP2 and WP3, data will be further processed at the premises of Nivel.

1.7 I am a member of a consortium of 2 or more partners. Clear arrangements have been made regarding data management and intellectual property. (also consider the possible effect of changes within the consortium on issues of data management and intellectual property)

- Yes, clear arrangements have been made regarding data management and intellectual property through a consortium agreement

The means and rights of consortium members to the data and software is delineated in a consortium agreement (Articles 9, 11 and Appendixes 2) and Nivel privacy regulation ‘Databank Communicatie in de Zorg’ Appendix 3 of the consortium agreement.

1.8 I can give an estimate of the size of the data collection; specifically, the number of participants or subjects (“n=”) in the collection and its size in GB/TB

- Yes (please specify)

In WP1: 5x20 audio-recordings of geriatric visits of 100 minutes each (100 GB) and geriatric visit reports from electronic patient files.
In WP2: 80 video-recordings of geriatric visits of 100 minutes each (80GB) and questionnaires.
In WP3: 15-20 Interviews from managers and practicing HPCs (general practitioners, internists).

Overall data collection size it is estimated to be ≈ 200 GB.

1.9 The following end products I will make available for further research and verification (please elaborate briefly)

- Syntaxes
- Other products (please specify)
- (Several versions of) processed data
- Raw data
- Data documentation

Where raw data pertains personal data, only processed (de-identified) data will be made available. Software documentation and partial code will be made available in an archive like Zenodo, supporting software citation (DOI). To increase findability and visibility we will use a software registry, the Research Software Directory (https://research-software-directory.org/).

1.10 During the project, I will have access to sufficient storage capacity and sites, and a backup of my data will be available. (please elaborate briefly)

- Yes, I will make use of my institution's standard facilities for storage and backup of my data

Questionnaire data are stored in the secured environment on the datacenter of PQR (IT service provider). Access is reserved for authorized employees only. The data backup runs daily. The daily back-ups are incremental and stored at 2 different locations, the original data is hosted in the data center of PQR. This guarantees long term storage of the data, as well as sufficient storage space.

The original video- and audio-recordings will be stored in the same secured environment, however, due to the sensitive nature of these files, these are only accessible from a PC on the premises of Nivel. Permission to access the files can only be granted with an agreement.

For the purpose of generating automated summarizations of medical visits, a technical programmer from Utrecht University will develop and implement a stand-alone transcription and pseudonymization service at Nivel. This transcription and pseudonymization application will be executed for the audio/video data by a Nivel project member at the Nivel building. Posteriorly, pseudonymized data (pseudonymized transcripts of recorded visits) will be shared with co-applicants at Utrecht University, Department of Informatics.

WP1

Utrecht University (Department of Informatics) will receive de-identified data (transcripts of audio-recordings and written reports of geriatric visits).

All data will be stored and preserved in Yoda. Yoda is Utrecht University's institutional research data repository, registered as such with re3data.org. Yoda complies with Utrecht University's Information Security policy for data classified as public, internal use or sensitive. Data along with its metadata will be shared within a closed user group, accessible to authorized users. All Yoda data are stored in at least two geographically spread locations. The data are stored and transmitted in encrypted format. Yoda complies with Utrecht University's Information Security policy for data classified as public, internal use, sensitive and critical.

Data transfer between Nivel and UU will be done via Yoda. A specific research folder will be created in Yoda environment for data sharing as Yoda enables collaboration with partners outside the UU.

2. Legislation (including privacy)

2.1 I will be doing research involving human subjects, and I am aware of and compliant with laws and regulations concerning privacy sensitive data.

- Yes, I will involve human subjects in my research. I will comply with the Algemene Verordening Gegevensbescherming (AVG)

- Wet op de Geneeskundige Behandelingsovereenkomst (Medical Treatments Contracts Act)

2.2 I will be doing research involving human subjects, and I have (a form of) informed consent from the participants for collecting their data.

- Yes, and this informed consent allows for the reuse of data (note that in the Code of Conduct for Medical Research, ‘reuse’ is also referred to as ‘further use’)

2.3 I will be doing research involving human subjects, and I will protect my data against misuse.

- Yes, the data will be pseudonymised. (please explain how this will be done, and by which organisation) and

Privacy by design will be incorporated in the project using both data-oriented and process-oriented strategies.
Only personal data connected to the research purpose will be collected.
Personal data will be stored securely at Nivel premises (see 1.10).
Nivel questionnaire data are stored in the secured environment of the datacenter of Nivel’s IT service provider. Access is reserved for authorized employees only. The original video-recordings will be stored in the same secured environment, however, due to the sensitive nature of these data, these are only accessible from a PC on the premises of Nivel. To this purpose, Nivel hosts a special, secured room where the privacy-sensitive video-recordings can be accessed and coded. Permission to access the recordings can only be granted after researchers, only those involved in the particular project, have signed a nondisclosure agreement.
A data steward and a statistician participate in the project to make sure regulations, storage and access are handled in the right way.
For WP1
Audio data will be transcribed, pseudonomyzed and masked using pseudonymization and masking software developed by the Department of Informatics at Utrecht University. The application of the transcription and pseudonymization software will be executed for the audio/video data by a Nivel project member at the Nivel building.
The pseudonymization key will be kept separately to the research data.
Personal data (pseudonymized transcripts of audio files and pseudonymized written reports) will be transferred securely between UU and Nivel using Yoda Repository, where data are stored and transmitted in encrypted format.
Yoda complies with Utrecht University's Information Security policy for data classified as public, internal use, sensitive and critical.
Access to these data will be controlled by the responsible researcher at UU via Yoda. Only restricted group members will have access to personal data.
A complete privacy risk assessment regarding the processing of personal data at Utrecht University in WP1 will be performed together with a data steward and privacy officer of the Faculty of Science at Utrecht University.

2.4 I will stick to the privacy regulations of my organisation

- Yes

3. Making data findable

3.1 The data collection of my project will be findable for subsequent research. E.g., on a catalogue, a web portal, or through the search enginge of the repository (note: this is key item 3, which you should report to ZonMw at the end of your project).

- Yes, it can be found through an online (metadata) catalogue or web portal (please specify)

Our research uses the research data management systems of Nivel for the storage and management of video recordings and questionnaires, and the Yoda system of Utrecht University (yoda.uu.nl) for the storage of pseudonymized transcripts and derived data.
WP 1
For WP1, Utrecht University will publish project metadata and de-identified, aggregated data, where possible, in Yoda repository (yoda.uu.nl), in compliance with the FAIR principles and will be made findable via its DOI persistent identifier.
Software developed at the UU will be developed on a git-based code repository. Software documentation and partial code will be made available in archive like Zenodo, supporting software citation (DOI) and to increase findability and visibility we will use a

software registry, the Research Software Directory (https://research-software-directory.org/).

3.2 I will use a metadata scheme for the description of my data collection (note: this is key item 7, which you should report to ZonMw at the end of your project).

- Yes, I will use a generic metadata scheme (please specify)

WP1
As a minimum, the metadata in Yoda will comply with the Datacite V4 standard. In addition, the system facilitates validation by per-research-configurable metadata schemas based on JSONschema to support discipline-specific metadata. The metadata are stored persistently along with the data and in the Yoda repository database. These are findable in Datacite upon publication of data. The metadata in the Yoda repository can also be retrieved via a service that complies with the Open Archives Initiatives Protocol for Metadata Harvesting (OAI-PMH).

3.3 I will be using a persistent identifier as a permanent link to my data collection (note: this is key item 1, which you should report to ZonMw at the end of your project).

- Yes, I will be using the DOI code

4. Making data accessible

4.1 Once the project has ended, my data will be accessible for further research and verification.

- Yes, immediately

4.2 Once the project has ended, my data collection will be publicly accessible, without any restrictions (open access).

- No, there will be access restrictions to my data collection (please explain)

Due to the characteristics of the collected data (personal data from patients, caregivers, health practitioners) only project metadata and de-identified, aggregated data (where possible) will be published, so that the data is not tracible to individual respondents.
Access to data will be done upon request and under restricted access.
Request for access to the data have to be approved by the steering committee.
The steering committee will take into account:
- the nature of the research
- the quality of the requiring researchers
- the nature of the requested data
- if the interest of applicants can be harmed
- the planning of applicants' own publications
Privacy sensitive data will not be shared.

4.3 I have a set of terms of use available to me, which I will use to define the requirements of access to my data collection once the project has ended (please provide a link or persistent identifier; also note that this is a key item 4, which you should report to ZonMw at the conclusion of your project).

- Yes, my institution has drafted a set of terms of use with the help of a legal advisor

- The application does not interfere with other current or planned activities
- The application does not harm the interests of the participants
- The requested data is suitable for answering the relevant research questions
- The results of the research must be publicly accessible
- The applicant must provide prior access to the draft publication, and allow all consortium partners 30 days for commenting on the publication
- At least one employee of consortium partners had to be actively involved in the publication

- Consortium partners and funder mentioned as source in the publications
- The end product will contain a short privacy section, provided by Nivel

A set of term for access to data collection is delineated in Article 10 of the collaboration agreement.

4.4 In the terms of use restricting access to my data, I have included at least the following:

- Conditions related to data security
- Agreements on methodology
- Collaboration in using the data set, including agreements on publication and authorship
- Whether or not the data set may be linked with another data set (for reasons of privacy)
- The manner in which the data set can be accessed
- The permitted period of use of the data set
- The approval of the participants allows for further research using this data set

5. Making data interoperable

5.1 I will select a data format, which will allow other researchers and their computers (machine actionable) to read my data collection (note: this is key item 5, which you should report to ZonMw at the end of your project).

- Yes (please specify)

As per table 1.3 data will be preserved (and where possible published) in preferred formats following international standards. Where the original data format is not a preferred format, data will be converted to a recommended (open) format following the recommendations from DANS (File formats | DANS (knaw.nl)).

5.2 I will select a terminology for recording my data (e.g., code, classification, ontology) that allows my dataset to be linked or integrated with other datasets (note: this is key item 6, which you should report to ZonMw at the conclusion of your project).

- Yes, metadata standard (please specify)

5.3 I will be doing research involving human subjects, and I have taken into account the reuse of data and the potential combination with other data sets when taking privacy protection measurements.

- Yes, the participants have given their permission for reuse of the data, and the data have been pseudonymised

6. Making data reusable

6.1 I will ensure that the data and their documentation will be of sufficient quality to allow other researchers to interpret and reuse them (in a replication package).

- I will perform quality checks on the data to ensure that they are complete, correct and consistent (please explain)
- I will document the software used in the course of the project (please specify)
- In addition, I will take further quality assurance measures (please specify)
- I will document the research process (please explain)

Nivel has a quality system and is ISO 9001 certified. Part of this system is the documentation of the research process by using a Logbook (journal) from the start, application, research methods, results of the report. We also perform quality checks on the data.
- Nivel employees work in accordance with internal work instructions to ensure quality of research. There are, for example,

instructions for taking interviews, data collection and data management.
- Nivel employees clean data in a structured way, the cleaning do-file is archived.
- There is a peer review process for all publications by colleagues.
- The data steward will prepare a data management plan.
- Nivel acts according to the Dutch code of science behavior.
- During the process, researchers will make use of the software MAXQDA to analyse qualitative data and STATA to analyse quantitative data. Code schemes (MAXQDA) and Syntaxes (STATA) will be archived.
- The audio-files and transcripts of the interviews will be archived, as well as the questionnaire, codebook, raw data and syntax of the questionnaires.

WP1: The source code will be stored in Yoda. Published code will be made available in public archive and software repository with accompanying documentation and a license clarifying terms of use.

6.2 I have a number of selection criteria, which will allow me to determine which part of the data should be preserved once the project has ended. (see also question 1.9 and 6.1)

- Some or all of the data must be destroyed once the project has ended, because of a contract or law

Data that contain privacy sensitive information will be stored separately, and destroyed after the project has ended apart from the recordings themselves which are stored at Nivel in conformity with Nivel regulations, without deadline.

6.3 Once the project has ended and the data have been selected, I can make an estimate of the size of the data collection (in GB/TB) to be preserved for long-term storage or archival.

- Yes (please specify)

In WP1: 5x20 audio-recordings of geriatric visits of 100 minutes each (100 GB) and geriatric visit reports from electronic patient files.
In WP2: 80 video-recordings of geriatric visits of 100 minutes each (80GB) and questionnaires.
In WP3: 15-20 interviews with managers and HCPs (general practitioners, internists).

Overall data collection size it is estimated to be ≈ 200 GB.

6.4 I will select an archive or repository for (certified) long-term archiving of my data collection once the project has ended. (note: this is a key item, which you should report to ZonMw at the conclusion of your project)

- Yes, and this archive has a data seal of approval (please specify the archive)

WP1
Utrecht University will use Yoda to archive and preserve research data. Metadata scheme in YODA complies with the Datacite V4 standard, supplemented by discipline specific metadata. The metadata are stored persistently along with the data in the Yoda repository database. Data or metadata (if data cannot be published) are findable, harvestable and searchable by Datacite upon publication of data or metadata. For unpublished archived data, an EPIC handle is available. Metadata are stored in machine readable formats. Yoda allows referencing of other datasets or publications that are part of the same publication or project and will show up in the metadata.

6.5 Once the project has ended, I will ensure that all data, software codes and research materials, published or unpublished, are managed and securely stored. Please specify the period of storage.

- Yes, in accordance with VNSU guidelines (please specify the number of years)

All software codes and research materials will be stored for at least 10 years after the end of the project. The audioand video-recordings of geriatric visits will be stored at Nivel without any deadline.

6.6 Data management costs during the project and preparations for archival can be included in the project budget. These costs are:

- Unknown (please explain)

Costs aren't specified for this project, but are covered in the standard working processes of Nivel projects.
For WP1, Utrecht University will cover the costs for data storage and preservation in Yoda.

6.7 The costs of archiving the data set once the project has ended are covered.

- Yes (please elaborate)

Costs aren't specified for this project, but are covered in the standard working processes of Nivel projects.
For WP1, Utrecht University will cover the costs for data storage and preservation in Yoda.