Project: POWER2DM
Grant Agreement No.: 689444
Call: H2020-PHC-2015
Topic: PHC-28-2015

D7.4: D7.2.1. Open research Data Management Plan

Workpackage	WP7
Task	T 7.2
Due date	31.07.2016
Submission date	31.07.2016
Deliverable lead	TNO
Version	1.0
Authors	A.A. de Graaf

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

1. Abbreviation

This section contains the abbreviations used in this deliverable.

Abbreviation	*Definition*
ADDQoL	**Audit of Diabetes DependentQoL**
ASQ	**After-Scenario Questionnaire**
BMI	**Body mass index**
DEPS-R	**Diabetes Eating Problem Survey-Revised**
D-FISQ	**Diabetes Fear of Injecting and Self-Testing Questionnaire**
DSMQ	**Diabetes Self-Management Questionnaire**
FCQ	**Fear of Complications Questionnaire**
GAD-7	**brief measure of Generalized Anxiety Disorder**
HbA1c	**Glycatedhemoglobin**
HFS	**HypoglycemiaFear Survey**
PAID	**ProblemAreas in Diabetes**
PHQ-9	**Patient Health Questionnaire**
PSS	**Perceived Stress Scale**
SMSS	**Self-Management Support System**
VAS	**Visual AnalogueScale**
WHO-5	**WHO-5 Well-Being Index**

2. Change procedure and history

This section contains the procedures for modifying the deliverable and maintaining a history of the changes.

Version	*Date*	*Changes*	*From*	*Review*
1.0	*August 4, 2016*			

Open Research Data Management Plan (DMP) for POWER2DM

1. Introduction

POWER2DM takes part in the Open Research Data Pilot on a voluntary basis:

https://www.openaire.eu/opendatapilot

and has included Article 29.3 in the Grant Agreement:

29.3 Open access to research data

For actions participating in the open Research Data Pilot: Regarding the digital research data generated in the action ('data'), the beneficiaries must:

(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;

(ii) other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan' (see Annex 1);

(b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39 of the Grant Agreement, all of which still apply.

As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in the DoW, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access.

The present Data Management Plan is created to outline how the research data collected and generated will be handled during and after the POWER2DM project. It will serve to make POWER2DM data FAIR:

Findable
Assign persistent IDs, provide rich metadata, register in a searchable resource,...

Accessible
Retrievable by their ID using a standard protocol, metadata remain accessible even if data aren't...

Interoperable
Use formal, broadly applicable languages, use standard vocabularies, qualified references...

Reusable
Rich, accurate metadata, clear licences, provenance, use of community standards...

According to principles outlined in www.force11.org/group/fairgroup/fairprinciples

The structure of the plan is as generated by the online tool DMP Online https://dmponline.dcc.ac.uk/ and the contents of the sections were drafted according to the guidance offered by DMP Online.

2. ADMINISTRATIVE DETAILS

Project Name: POWER2DM - Predictive model-based decision support for diabetes patient empowerment

Project Identifier: POWER2DM

Grant Title: NUMBER — 689444 — POWER2DM

Principal Investigator / Researcher: Albert A. de Graaf (Coordinator)

Description: Data Management Plan for POWER2DM Health and Observations of Daily Life data used for self-management by diabetes patients

Funder: European Commission (Horizon 2020) Call Topic: PHC 28 – 2015: Self-management of health and disease and decision support systems based on predictive computer modelling used by the patient him or herself

3. DATA SETS

A wide range of patients with diabetes might benefit from support through POWER2DM. Two different patient populations with altered glucose metabolism are targeted in POWER2DM: T1DM and T2DM in primary/secondary & tertiary care. The Self Management Support System (SMSS) developed in POWER2DM will be tested in a pragmatic RCT: the **POWER2DM Evaluation Campaign**, to establish the accuracy and utility of the POWER2DM DSS and APIs and evaluate effectiveness in a real-world setting. We will use three different centres who are specialized on (some or all of) these entities: The Reina Sofia University Hospital in Spain (T1DM), the Leiden University Medical Centre and Primary

Care Research Network in the Netherlands (T1&2DM), and the Institut für Diabetes "Gerhardt Katsch" in Karlsburg, Germany (T1&2 DM). Study characteristics are as follows:

Study design and Operation: The protocol for the POWER2DM Evaluation Campaign will be pragmatic randomised trial with 9 months follow-up of individual patients. Patients will be randomised to either Power2DM support (active arm) or usual care (control arm). Patients in the Power2DM intervention the first 2 weeks patient will follow an established protocol in order to monitor any problems in using the whole system. There will be evaluation moments at baseline, after 3, 6 and 9 months.

Endpoints: Primary outcome: %Hba1c levels before and after the intervention between the two arms (active versus control). Secondary outcomes: Generic Quality of Life (SF-36); Patient utilities (EQ5D-L); Disease Specific Quality of Life (DSQLS); costs (CostQ); self-management outcome (heiQ: Health Education Impact Questionnaire, Summary of Diabetes Self Care Activities, Diabetes Management Self-efficacy); lifestyle and physical activity and other process outcomes of the POWER2DM modules and services for patients and care providers: reliability, usability, acceptance and actual usage.

Sample Size: Variable: The level of Hba1c%. Minimum detectable difference: 0.35% (Standard Deviation 1.0%). For an alpha error of 0.05 and a power of 80%, the minimum sample size needed is 129 subjects per group. **The POWER2DM RCT will include 140 type 1 DM and 140 type 2 DM subjects, 280 patients in total**, allowing us to face a loss to follow of 8.5%. In pre-specified subgroup analyses of patients with T1DM and T2DM we are able to detect a difference of 0.5% with a sample size of 63 subjects per treatment strategy per DM subtype (N=70 with 10% loss to follow-up).

Statistical analysis: The primary outcome will be analysed using the Stata 13 xtmixed command for multi-level linear regression, adjusting for clusters at GP-level, repeated measurements within a patient values (StataCorp, College Station, Tx, USA). Strategy by time interactions will be assessed to detect differences between the groups at particular time points. In addition, strategy by time by DM type will be assessed to detect differences in effects between the two DM subtypes.

Data from participants will be recorded in the POWER2DM Personal Data Store. Appropriate privacy and data security measures will be put in place acoording to pertinent regulations. Personal data will be stored with an anonymised identifier. The keys that will enable to link data to an individual person will be stored in a secured fashion on separate servers. For the Open Research data pilot, part of the contents of the Personal Data Store will be made available for research purposes and transferred to a Data repository according to cinformed onsent provided by study participants.

Comment [Ad1]: Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.
Questions to consider:
- What data will you create?

Guidance:
Give a brief description of the data that will be created, noting its content and coverage

4. Data set description

POWER2DM will collect basic diabetes related data, clinical measurements, patient data (Quality of Life-QoL questionnaires, self-management profile), daily nutritional intake, exercise level, sleep quality, glucose measurements, vital signs (pulse, temperature), medication intakes, etc. The provisional list of measured parameters subdivided in 4 categories is as follows:

Comment [Ad2]: We may wish to add a 5th category containing model predictions, and a 6th category containing use characteristics of the various devices and POWER2DM features

Table 1. Provisional list of measures in POWER2DM Evaluation Campaign dataset:

Measure Category and Name (# items)	Code
Lifestyle and Daily Monitoring	**LDM**
Blood Glucose Level	1
Dietary Intake	2
Activity Tracker	3
Sleep Tracker	4
Sleep Quality VAS	5
Relaxation/Stress	6
Stress VAS (1)	7
Emotional VAS (1)	8
Diabetes Medication Treatment (Type/ Dosage/Frequency)	9
Questionnaires (# items)	**Q**
WHO-5 (5)	1
PHQ-9 (9)	2
GAD-7 (7)	3
PSS (10)	4
PAID (20)	5
DSMQ-R (20)	6
ADDQoL (28)	7
HFS (27)*	8
DEPS-R (14)*	9
FCQ (15)*	10
D-FISQ (21)*	6
Gut Health (?)	11
ASQ (3)	12
Clinical/Lab Tests	**CLT**
HbA1c	1
Fasting Glucose	2
Fasting insulin	3
Insulin sensitivity (%HOMA-2 S): based on fasting glucose/insulin	4
Beta cell function (%HOMA-2-B): based on fasting glucose/insulin	5
Inflammation (mg/l hs-CRP)	6
Tissue damage (TC, HDL-C, LDL-C,TG, liver damage blood markers, kidney damage markers, neuropathy markers, smoking status)	7
Non-estrefied fatty acids	8
PatientCharacteristics	**PC**
Anamnese: Age/ Gender/Height/Type Diabetes /MedicalHistory (Time since diagnosis and Complications)/AS4	1

Weight	2
BMI (calculated from Weight and Height)	3
Waist	4
Blood pressure	5

Note: *indicates that this measure will only be used if a patient engages in an associated self-management task (e.g. only insulin users will be asked about anxiety related to using insulin) or they indicate associated problems in other questionnaires (e.g. DEPS-R will be administered if the patient indicates issues regarding eating)

5. Data Capture methods:

The data of the different categories in Table 1 will be captured as follows:

- How will the data be created?

The following data sources are planned:

- PC category: measured by medical professional
- CLT category: measured in Clinical Chemistry lab
- Q category: patient self-evaluation
- LDM category: data will be registered by devices and sensors used by the patients, or entered by the patient in a mobile or web application

Comment [Ad3]: We still need to decide: 3separate local labs in the 3 pilot regions or a single central lab?

Comment [Ad4]: We need to decide whether filling out of questionnaires will be supervised or not

- What standards or methodologies will you use?

The following standards will be used:

- PC category: medical professional standards
- CLT category: Clinical Chemistry lab methods/standards
- Q category: established questionnaire methods/standards/scoring methodology;
- LDM category: standards and methods acoording to specifications of devices and sensors used

- How will you structure and name your folders and files?

Folders will be named according to pilot site name / measure category/measure code (cf. Table 1).

File names will be named according to the (anonymous) subject identifier. The naming will be Subject_Site_Category_Code. Each entry in a file specifies a value plus the associated date/time specification.

Comment [Ad5]: This is a suggestion. It will result in a large set of files (~30 per patient)

- How will you ensure that different versions of a dataset are easily identifiable?

We do not at present envision different versions of the dataset since only unprocessed data are collected.

Comment [Ad6]: open for discussion. We may want to store some processed data as well e.g. weekly aggregates of physical activity , calorie intake,sleep, stress, etc.

6. Metadata:

- How will you capture / create the metadata?

We plan to capture the metadata in a "readme" text file. The strict file naming convention should further allow to uniquely identify which data is contained in each file.

Comment [Ad7]: open for discussion

- Can any of this information be created automatically?

The "readme" text file will be created by hand. The data file names will be created automatically.

- What metadata standards will you use and why?

POWER2DM will create a dataset of a diverse nature, not matching any of the disciplinary metadata standards for Biology, Earth Science, Physical Science, Social Science & Humanities, and General research Data offered by DCC on their website http://www.dcc.ac.uk/resources/metadata-standards.

We do not consider the development of a dedicated standard for POWER2DM essential for the efficient dissemination of the project results. The proposed metadata capturing, based on the documentation in the "readme" text file together with the strict file naming convention will allow interested researchers to re-use the data without much difficulty.

7. Data sharing, repository and restrictions

The POWER2DM dataset concerns personal medical and behavioural data. Data storage in the POWER2DM Personal Data Store will be subject to strict privacy/security measures dictated by the Ethics criteria that apply to the project (Ethics Deliverables of Work package 9). In transferring data to a data repository for sharing, special care will be taken to preserve the same standard of data privacy/security. This will be accomplished by properly anonymising the data and ensuring that the keys to link data to patient identity are not transferred. As an additional privacy precaution, data may be aggregated to a certain extent depending on requirements of the POWER2DM models (i.e. still allowing the reproduction of the results).

As a guiding principle for sharing, study participants are considered owner of their personal data. Therefore, participants will be asked to participate in the Open Research Data Pilot by giving informed consent to their data being made publicly available after proper anonymisation and aggregation mentioned above. This consent will be asked in a second separate consent form, in addition to the standard informed consent to have their data made available to the project team for research purposes. While the latter is required for participation in the study, the response to the Open Research Data Pilot sharing consent form will not be part of the inclusion criteria.

Method for data sharing:

- How will you make the data available to others?

The data will be stored in a data repository

- With whom will you share the data, and under what conditions?

The data will be publicly available for any party without the requirement to attribute the data to the POWER2DM Consortium (Open Access, Creative Commons CC Zero License (cc-zero) (see http://ufal.github.io/public-license-selector/)

Comment [Ad8]: proposed

- Are any restrictions on data sharing required? e.g. limits on who can use the data, when and for what purpose.

No restrictions on who can use the data and for what purpose apply.

- What restrictions are needed and why?

An embargo period of maximum 12 months after finalization of the project is deemed required to allow sufficient time for publication of the results, and for establishment of intellectual property (patent applications).

Comment [Ad9]: suggestion

- What action will you take to overcome or minimise restrictions?

Subjects participating in the study will be asked to give separate informed consent to make their data publicly available for any purpose. Publications and patent applications will be planned as early as possible yet realistic.

- Where (i.e. in which repository) will the data be deposited?

The data plus instruction files for usage will be deposited in the Zenodo cost-free data repository for sharing (http://www.zenodo.org). The Zenodo procedures for long-term preservation of the data will be put in place. The duration of data preservation is still to be decided. The data is not of a very complex nature. The approximated end volume will depend on several factors including the number of participants willing to take part in the Open Research Data Pilot, and the degree of aggregation to be applied, and as a consequence is difficult to predict at the current time.

No associated costs will be involved with the data sharing.

8. Preservation Plan:

The following applies:

- What is the long-term preservation plan for the dataset?

The dataset will be deposited in the Zenodo data repository.

- Will additional resources be needed to prepare data for deposit or meet charges from data repositories?

No. The data preparation for deposit is part of Workpackage 7 (Dissemination) and the deposit in Zenodo is free of charge. If psossible, the physical depositing of the data will be done already before the end of the project, but the embargo will be in place until the end of the period required for publications and securing of intellectual property.

- What additional resources are needed to deliver your plan?

No additional resources are needed. Since the data will be publicly available without restrictions, we do not need to keep a supervised data release system in place.

- Is additional specialist expertise (or training for existing staff) required?

No additional specialist expertise is required. The readme files supplied with the deposited data will contain all the information required to use the data.

- Do you have sufficient storage and equipment or do you need to cost in more?

This does not apply since the dat will be deposited in the Zenodo repository

- Will charges be applied by data repositories?

No. Zenodo is a cost-free repository.

- Have you costed in time and effort to prepare the data for sharing / preservation?

Yes. This is part of Workpackage 7 Dissemination.