



DR 1.1: Initial Data Management Plan

Rosemarijn Looije*, Elettra Oleari[†]

**TNO, Soesterberg, The Netherlands*

[†]FCSR, Milan, Italy

<rosemarijn.looijs@tno.nl>

<i>Project, project Id:</i>	EU H2020 PAL / PHC-643783
<i>Project start date:</i>	March 1 2015 (48 months)
<i>Due date of deliverable:</i>	August 31 2015
<i>Actual submission date:</i>	August 3 2015
<i>Lead partner:</i>	TNO
<i>Revision:</i>	final
<i>Dissemination level:</i>	PU

This deliverable presents the initial data management plan (DMP) for the PAL project.

1	Initial DMP	4
1.1	Data set description	4
1.2	Standards and metadata	7
1.3	Data sharing	7
1.4	Archiving and preservation (including storage and backup)	8

Executive Summary

The key objective of the Data Management Plan (DMP) is to ensure well-managed and well preserved data. It considers data gathering as well as metadata generation, data analysis and data storage. This initial DMP is created in DMPonline <https://dmponline.dcc.ac.uk>.

Role of DMP in PAL

The DMP contributes in reaching the overall goal of PAL by specifying the data needed for user and functional requirements as well as the data needed to evaluate the effects of PAL on child, informal and formal caregiver. It also supports structured data gathering and storage.

Contribution to the PAL scenarios and prototypes

The DMP supports an efficient gathering of information and by continuous refinement after interaction with the stakeholders it guides the development of scenarios that are able to evoke effects from PAL.

1 Initial DMP

1.1 Data set description

To reach our objective of designing, implementing and evaluating the PAL system with stakeholders involved in each phase, we will apply different research methods. The user and functional requirements are derived in a continuous process during the project in which we apply co-design/co-creation techniques during interactions with the stakeholders (formal and informal caregivers and children). Structured interviews, focus groups, observational studies and questionnaires will be used to derive user and functional requirements. Furthermore by doing fast prototyping stakeholders can provide their input on concrete aspects of the system to improve implementation. Finally during more formal evaluations we will make use of questionnaires, logging and observational data which is analyzed both statistically and more ethnographically (e.g. grounded theory) to evaluate the system and provide refined user requirements.

Below we provide a table (see Table 1) in which all data is structured: the "what" column tells the kind of data collected, the "data storage" column describes what type is stored and the "form" column describes how it is stored. Not all data will always be analyzed to result in all suggested data forms. At this moment we have little information on which datasets we are going to link with each other to provide other relevant metadata sets. This will be further addressed in the next version of the DMP.

Data gathered within the FP7 project ALIZ-e will be used as starting point and new information will be integrated.

What	Data storage	Form
Consent forms (Parents/custodian, children w and w/o T1DM, formal caregivers, teachers)	paper, scanned	PDF
Participant data (gender, age, years T1DM, robot experience, ehealth experience, type of diabetes therapy - insulin pump, multi-injective)	Paper/(excel) table	xlsx/csv
Requirement elicitation with Structured interviews	Video/Audio Recordings	docx/pdf, mp4/mov/mpg/wmv, mp3/wav
Requirement elicitation with Focus groups	Video/Audio Recordings, observation notes, output (e.g. drawings)	Mind maps, docx/pdf, mp4/mov/mpg/wmv, mp3/wav
Requirement elicitation with Observational studies	(Video/Audio Recordings), observation notes	Observer forms with recurring aspects (docx, xlsx), mp4/mov/mpg/wmv, mp3/wav, docx/pdf
Questionnaires on user requirements	Paper/computer	docx/xlsx/csv
Performance data	logging	xlsx/csv
Adherence	logging, data input	xlsx/csv, docx/pdf
Emotional state	data logging, questionnaires, photos (selfies), observation notes, dialogue data (speech/text), video	xlsx/csv, jpg/png, mp4/mov/mpg/wmv, mp3/wav, docx/pdf
PAL experience (child, formal and informal caregivers)	questionnaires, observations notes	xlsx/csv, docx/pdf
autonomy/relatedness/competence feelings	questionnaires, observations notes, dialogue data (speech/text)	docx/xlsx/csv, mp4/mov/mpg/wmv, mp3/wav, docx/pdf
Glucose values, nutritional and lifestyle habits	logging, explicit input user	xlsx/csv, database

What	Data storage	Form
parent/custodian questionnaires on parenting capacities that influence disease management (pre/post, compared to parents of healthy children) - e.g parental overprotection, perceived child vulnerability, parenting stress	questionnaires	docs/xlsx/csv
parent/custodian and/or teacher questionnaires (attitude/knowledge/trust/skills/shared responsibility)	questionnaires	docx/xlsx/csv
Professional care-giver questionnaires (Trust/acceptance/awareness/tailoring/PAL experience/effect on child)	questionnaires	docx/xlsx/csv
User, functional and design requirements	Derived from all data	sCET database, xml, word or excel

Table 1: Data structure

1.2 Standards and metadata

For transcription we will use word (.docx) or special transcription software (to be decided on). For mind maps we will use powerpoint (.pptx). Data will be represented in Excel (.xlsx or .csv), data analysis output will be in SPSS, PRISM or R for quantitative analysis and with atlas.ti or word for more qualitative observation analysis. Recordings are saved in mp3 (audio), mp4 (video) or another widespread format.

Use cases, requirements and claims are stored in the situated Cognitive Engineering format (online) which will be exported to doc, html or xml.

All data will be supported by an ontology in RDF format.

All data will be collected in folders using the following format YEAR_MM_partnerAcronym_location_experiment. The DMP is further supported by an experimentation report template in docx format with information about the experiment performed (main researcher, goal, lessons learned, time of execution, partners involved, methodology summary, overview of data outcomes (references to data storage), conclusions and references to publications.

1.3 Data sharing

Most data will be anonymous and therefore free accessible for the scientific community, recordings of voice and face are ethically a different issue as might be the case with medical data, some dialogues and connections between data sets (e.g. glucose values and diary input). The data that is provided, disclosed or otherwise made free accessible shall not include personal data as defined by Article 2, Section (a) of the Data Protection Directive (95/46/EEC).

Particular attention will be taken for video and some of the other data. In this case we will take all necessary steps in order to ensure that the data and video will be accessible only after the signature of a specific written agreement that imposes that the same data and video will not be shared. The modalities and possibilities to sharing data and video will depend on the written informed consent given by caretakers and children and by the ethical committees of the partners involved. For sharing data and video between partners we have a specific Material Transfer Agreement (MTA), which can also be used outside the project if we want to share outside of the consortium. In this specific case the MTA will be modified considering that the recipient parties are not partners of the consortium.

As far as possible and useful for the community we will put the data on the OpenAire supported Zenodo repository (<https://zenodo.org/>).

1.4 Archiving and preservation (including storage and backup)

All data is preserved either in the Zenodo repository, the project SVN at TNO or for more sensitive data at the specific partners locations (e.g. the medical data connected to the user). In particular, data collected during the field experiments with the stakeholders will be stored and preserved in each of the leading country partners: Netherlands and Italy.

This data will be preserved according to the rules of research data, which is at least 5 years after the project.