Deliverable

D1.7. Data Management Plan: first update

Workpackage	WP1
Lead beneficiary	CEA
Due date	31st of July 2018
Submission date	03/09/2018
Type	PU
Author(s)	Margarita Anastassova, Sabrina Panëels, Diana Dimitrova, Francesca Pichierri
Abstract	This document is the second iteration of the data management plan, which describes what data the project will generate, how they will be produced and analysed. This updated version presents a new classification of the type of data collected and shared during the project.
Keywords	

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 689947.

History of changes

Rev. N	Description	Author(s)	Date
1	First version	Margarita Anastassova	03/06/2018
2	Data to be included	Diana Dimitrova, Francesca Pichierri	15/06/2018
3	Second version	Margarita Anastassova	05/07/2018
4	Update from partners on data to be shared	All	24/08/2018
5	Final version	Sabrina Panëels	01/09/2018

Table of contents

Deliverable Description

This document is a deliverable of the STARR project, which is funded by the European Union's Horizon 2020 Programme under Grant Agreement #689947. It describes what data the project will generate, how they will be produced and analysed. It also aims to detail how the data related to the STARR project will be disseminated and afterwards shared and preserved.

1. Scope of the document

The STARR project Data Management Plan (DMP) primarily lists the different datasets that will be produced by the project, the main exploitation perspectives for each of those datasets, and the major management principles the project will implement to handle those datasets.

The purpose of the DMP is to provide an analysis of the main elements of the data management policy that will be used by the consortium with regard to all the datasets that will be generated by the project.

The DMP is not a fixed document. On the contrary, it will have to evolve during the lifespan of the project. This second version of the DMP includes an updated overview of the datasets to be produced by the project, and the specific conditions that are attached to them. In particular, it provides a new classification of the type of data. Details about the privacy aspects will be included in the deliverable D8.3 due at the end of September 2018. The final version of the DMP will in addition describe how the system will use data when the development is finished and it is running semi-autonomously. Thus, the data management plan will cover all the data life cycle.

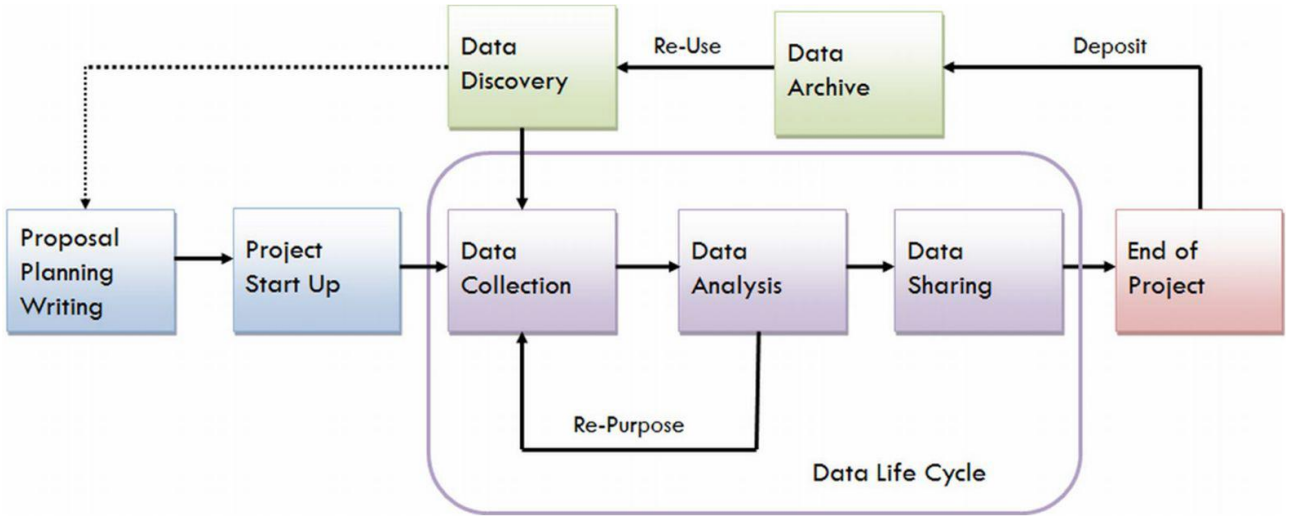


Figure 1: Steps in the data life cycle. Source: From University of Virginia Library, Research Data Services

2. Information about the project

The table below provides synthetic information about the STARR project.

Name	Decision **S**uppor**T** and self-m**A**nagement system for st**R**oke survivo**R**s
Acronym	STARR
Project objectives	It is crucial to provide stroke survivors with information, problem-solving skills, motivational levers, confidence, and continuous professional and peer support to help them better manage their condition, live independently, and prevent a secondary stroke. At the same time, it is also crucial to inform, with the stroke survivor's permission, their carers and the medical staff about the evolution of the stroke survivor's lifestyle and the risk of stroke. This is the primary objective of the STARR project. We will develop a modular, affordable, and easy-to-use, install and maintain system, which will inform stroke survivors about the relation between their daily activities (e.g., medication intake, physical and cognitive exercise, diet, social contacts) and the risk of having a secondary stroke.
Keywords	Stroke, prevention, risk factors, monitoring, daily activities, sensors, self-management
Call	PHC-28-2015
Funding body	European Commission, H2020 program
Grant agreement No	689947
Members of the consortium	**Coordinator**: French Atomic and Alternative Energies Commission, LIST Institute (CEA LIST) **Partners**: Bluelinea SA, Fondation Hopale, The Stroke Association, RT-RK, InitHealth, Lunds Universitet, University of Luxembourg, FIZ Karlsruhe – Leibniz-Institute for Information Infrastructure, Osakidetza-Servicio Vasco de Salud, Magillem Design Services
Contact	Margarita Anastassova margarita.anastassova@cea.fr

Contact's affiliation	CEA LIST, Sensory and Ambient Interfaces Laboratory
Start date of the project	01/02/2016
Duration	42 months

3. Responsability for the data

Person in charge of the data management during the oroject	Margarita Anastassova (personal and non-personal data) margarita.anastassova@cea.fr Franziska Boehm (specific focus on EU regulations about personal data management) franziska.boehm@kit.edu
Partners' reponsability	Every partner is responsible for the data they are collecting. This is valid for both personal and non-personal data. The data will be collected, combined, stored and transmitted according to the relevant national, European and institutional regulations. As far as the management of personal data is concerned, the partners will be supported by FIZ Karlsruhe, provinding guidance on personal data management in an EU perspective.
Data management policy	All personal data collection in STARR will be done within the remit of formal ethics clearances obtained at our testing sites and granted by the relevant university and/or local health officials. Thus, any patient-related data, such as data from pre-exiting health record data and behavioural data depicted, for example, by keyboard presses, video, audio, motion tracking will fall under the ethics clearance. Some non-patient data, such as requirements from stakeholders in WP3, in the form of anonymous questionnaire responses & focus group opinions, can be gathered more informally. The legal basis for the personal data processing will be the participant's consent, obtained in accordance with the rules to which the collecting partner is subject. The most relevant standards regarding data handling, in this experimental context with patients, concern the area of ethics, data protection and privacy. They are listed below:

	- Directive 95/46/EC: Protection of individuals with regard to the processing of personal data and on the free movement of such data. - Directive 99/5/EC: Radio equipment and telecommunications terminal equipment and the mutual recognition of their conformity. - Directive 98/44/EC: Legal protection of biotechnological inventions. - Art. 7, 8 Charter of Fundamental Rights - Art. 8 European Convention on Human Rights - Case Law by the European Court of Human Rights - Documents of the Aricle 20 Working Party - Ilves report on E-health - GDPR regulation (2016/679). - Draft data protection regulation. Ad-hoc behavioural data that we will collect will be of a new form and so no standards yet apply to its usage. In the wider sense of knowledge, management standards and good practice, we will of course record and learn from our activities in the first cycle of STARR to inform our practice from then on. All of our patient data will be reported as anonymized group summaries in the project deliverables and in peer reviewed publications. Anonymous individual quotes describing requirements will also be used in these reports. During system evaluation, user performance data (e.g. task execution time and number of errors) is transient and context sensitive (to the particular task, sensors and user). There is no public value in the data and hence no foreseen need for public access beyond that of the therapist. Nonetheless we have no objection in principle to releasing this data in the spirit of Open Access if requested.
Ownership and access to data	A consortium agreement was negotiated and signed by all the parties in order to inter alia specify the terms and conditions pertaining to ownership, access rights, exploitation of background and results and dissemination of results, in compliance with the grant agreement and Regulation n°1290/2013 of December 11th, 2013. The consortium agreement was based on the DESCA Horizon

2020 Model Consortium Agreement with the necessary adaptations considering the specific context and the parties involved in the project. Its basic principles are as follows:

- The parties will exhaustively identify the background intellectual property they will bring to the project, and assess its availability for access rights as regards to potential third parties' rights over such background;

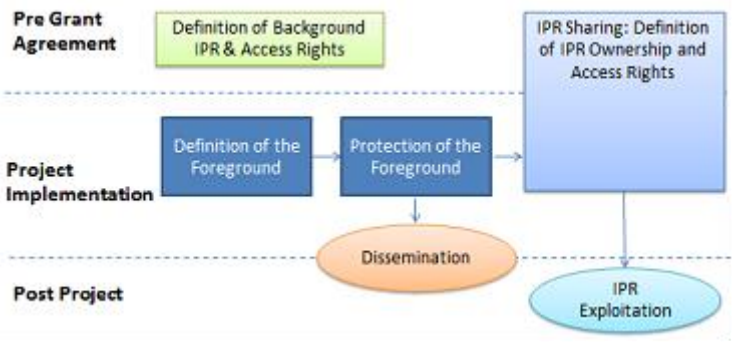
- Ownership of results including joint results generated by two or more parties will go to the party(ies) having generated such results;

- The owning parties will take all appropriate measures for the protection of the results capable of commercial or industrial exploitation, notably through intellectual property rights when relevant;

- The parties will use their best efforts to exploit and disseminate the results, either directly or indirectly, for instance by out-licensing said results;

- Each party will give access rights (through licenses) to their background and results to the other parties for the implementation of the project and/or for the exploitation of those other parties' own results (under fair and reasonable conditions).

Knowledge management follows the strategy presented in the Figure below. The IPR activities are organised according to the different project phases.



In principle, Foreground is managed according to the provisions of the European Commission, and the access to the foreground created throughout the project lifetime is specified by the Consortium Agreement. As a general rule, the foreground is considered as a property of the Contractor

	generating it, and in this sense the originator is entitled to use and to license such right without any financial compensation to or the consent of the other Contributors. In case of licensing to third parties, the Contributors shall be informed in advance and appropriate financial compensation shall be given to them. Starting from these basic rules, other particular situations will be treated in WP9 (Dissemination & Exploitation): - If the features of a joint invention, design or work are such that it is not possible to separate them, the Contributors could agree that they may jointly apply to obtain and/or maintain the relevant rights and shall strive to set up a joint exploitation agreements in order to do so; - An originator of the foreground could decide not to seek protection of its Foreground. In this case, another contractor interested in such protection might apply for it, advising the other Contractors. In case several Contractors are interested, an agreement is necessary between them. Concerning access rights, each Contractor shall take appropriate measures to ensure that it can grant Access Rights and fulfill the obligations under the EU Contract. The Contractors will agree that Access Rights are granted on a non-exclusive basis, and that, if not otherwise provided in the Consortium Agreement or granted by the owner of the Foreground or Background, the Access Rights does not include the right to grant sub-licenses. The Consortium agreement will dedicate one section or one appendix to define which access rights to the background may be granted. Access rights to foreground and background needed for the execution of the project will be granted on a royalty-free basis to the project partners. Publication and dissemination of foreground are granted with the approval of the Consortium, making sure that the period of secrecy needed for IP protection is respected. Contractors have to inform the Consortium and the Commission of its intention to publish on its foreground. Adequate references to the Contract with the EC shall be given in the publication. Publication can be impeded if another contractor can show that the secrecy of the foreground is not guaranteed.

Publication of research data	Since in STARR certain research activities rely on the processing of personal data of stroke survivors, the privacy and the protection of their personal data should be ensured. However, this does not exclude the publication of aggregated, properly anonymized data, e.g. in statistical form. In this way both the privacy of the stroke survivors will be respected and scientific results can be disseminated

4. Data set description

The project partners have identified the dataset that will be produced during different phases of the project. The list is provided below, while the nature and details for each dataset are given in the subsequent sections. This list is an adapted version of the first deliverable with a new holistic classification, and will be further refined for the last iteration of the DMP.

Data set name	**Personal identifiable data**
Type of data	Name, gender, age, marital status, deprivation index, hand dominance, educational level, type of job, hobbies, socio-familiar support
Format	Documents (paper and digital such as Word documents or Excel sheets), pictures, audio or video recordings, inputs to the STARR application
Source	This data comes from: - Questionnaires (paper originals, responses captured in an Excel spreadsheet), analysis (e.g. charts) within the spreadsheet - Interviews transcribed to Word documents (anonymised), analysed data from the interviews - Focus groups: discussions transcribed to Word documents (anonymised), synthesis of themes captured in Word documents - Consent forms: signed paper documents - Inputs to the application - Video or audio recordings of user testing and studies - Pictures from user testing and studies
Reuse and sharing	Only relevant anonymised and aggregated data are transmitted and reused by the partners not collecting the data.
Archiving and preservation (including storage and backup)	The data will be stored by the partner collecting it (on their own computers and/or institutional servers).

Data set name	Sensitive data (except Health data)
Type of data	Ethnic group, sex life or orientation, political opinions, religious beliefs or other beliefs of a similar nature
Format	Documents (paper and digital such as Word documents or Excel sheets), audio recordings or videos, inputs to the STARR application
Source	This data comes from: - Questionnaires (paper originals, responses captured in an Excel spreadsheet), analysis (e.g. charts) within the spreadsheet - Interviews transcribed to Word documents (anonymised), analysed data from the interviews - Focus groups: discussions transcribed to Word documents (anonymised), synthesis of themes captured in Word documents - Inputs to the application - Audios or videos recordings of user testing and studies
Reuse and sharing	Only anonymised and aggregated relevant data are transmitted and reused by the partners not collecting the data.
Archiving and preservation (including storage and backup)	The data will be stored by the partner collecting it (on their own computers and/or institutional servers).

Concerning sensitive data, it also includes health data. However, we chose to separate it and create a specific category as this is the major type of data that will be collected and analysed during the project. It is further detailed in the table below.

Data set name	Health data (physical + mental)
Type of data	Medical history, blood pressure, lipidic profile, glycaemia, heart rate, diet, toxic habits, disabilities, weight, depression, stress, pain, motor function, physical activity, canes and wheel chair use, gait analysis, skeleton movements in the game, emotions, behavior, adherence to treatment, information needed and obtained from the psychological model analysis

Format	Documents (paper and digital such as Word documents or Excel sheets), videos, inputs to the STARR application, visual sensing (Kinect or equivalent)
Source	This data comes from: - Questionnaires (paper originals, responses captured in an Excel spreadsheet), analysis (e.g. charts) within the spreadsheet - Interviews / Focus groups transcribed to Word documents (anonymised), analysed data from the interviews - Inputs to the application via questions or sensing - Video or audio recordings of user testing and studies - Movement recordings by the games and training applications - Movement recordings by the wearables
Reuse and sharing	Only relevant anonymised and aggregated data are transmitted and reused by the partners not collecting the data.
Archiving and preservation (including storage and backup)	The data will be stored by the partner collecting it (on their own computers and/or institutional servers).

5. Metadata

Metadata is data on the research data themselves. It enables other researchers to find data in an online repository and is, as such, essential for the reusability of the dataset. By adding rich and detailed metadata, other researchers, can better determine whether the dataset is relevant and useful for their own research. In the online depositories used by STARR partners, metadata (type of data, location, population etc.) will be uploaded in a standardized form. This metadata will be kept separate from the original raw research data.

6. Data volume

The volume of data collected by the STARR platform will be limited to the time during which the pilot will be run by each patient, as well as to the total number of participating patients and the quantity of information that will be sent from the client agents to the platform. It is estimated that the total amount of information gathered from telemonitoring without including the smart-space will be under 100Gbytes making the following assumptions (N = 20, effective time pilot = 2 months, daily system usage = 12 hours, sensors data rate = 1Kbytes/s, which estimate a total of 52Gbytes). This information will be duly protected as described in section 7.

7. Data security

The STARR project will use methods that emphasize good field access and extended contact and trust building with participants. Due to the sensitive nature of some of the topics that will be discussed in interviews and focus groups, data security is of vital importance. The following guidelines will be followed in order to ensure the security of the data:

- Keep anonymised data and personal data of respondents separate;
- Encrypt data if it is deemed necessary by the local researchers;
- Store data in at least two separate locations to avoid loss of data;
- Do not store personal data on USB drives;
- Save digital files in one the preferred formats (see table above), and
- Label files in a systematically structured way in order to ensure the coherence of the final dataset.
- Whenever possible and according to the ICT policies of each organisation, on all devices (desktops, laptops, external hard-disks, USB drives, etc.) that will be used for archiving, storing and data transferring will be encrypted.