Serums

Project no. 826278

SERUMS

Research & Innovation Action (RIA)
SECURING MEDICAL DATA IN SMART PATIENT-CENTRIC HEALTHCARE SYSTEMS

Data Management Plan
D8.3

Due date of deliverable: 30th June 2019

Start date of project: January 1st, 2019

Type: Deliverable
WP number: WP8

Responsible institution: University of St Andrews
Editor and editor's address: Vladimir Janjic, University of St Andrews

Version 1.0

Project co-funded by the European Commission within the Horizon 2020 Programme		
Dissemination Level		
PU	Public	√
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)	

Change Log

Rev.	Date	Who	Site	What
1	21/06/19	Vladimir Janjic	USTAN	Initial Version

Executive Summary

This deliverable describes the **Serums** project Data Management Plan (DMP) that specifies the management and access rights of the data used to produce the research output of the project.

Contents

1. Introduction

This deliverable describes the Data Management Plan for the **Serums** project. The Horizon 2020 programme requires all projects to adhere to the conditions of the Open Research Data Pilot [1]. Beneficiaries participating in the Pilot are required to:

- Develop (and keep up-to-date) a Data Management Plan (DMP).

- Deposit the data in a research data repository.

- Ensure third parties can freely access, mine, exploit, reproduce and disseminate the data.

- Provide related information and identify (or provide) the tools needed to use the raw data to validate the research.

The pilot applies to

- The data (and metadata) needed to validate results in scientific publications.

- Other curated and/or raw data (and metadata) that you specify in the DMP.

It is important to note that beneficiaries participating in the Pilot are not obliged to make all datasets open (as described in their DMP, and in compliance with confidentiality, security, data protection, and other relevant considerations). In addition, the DMP is intended to be a live document, that will be updated throughout the project to reflect on possible restrictions on new sets of data produced/used by the project.

---

[1] https://www.openaire.eu/opendatapilot

2. Data Management Plan

The aim of a Data Management Plans (DMP) is to describe the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP includes information on:

- the handling of research data during and after the end of the project

- what data will be collected, processed and/or generated

- which methodology and standards will be applied

- whether data will be shared/made open access and

- how data will be curated and preserved (including after the end of the project).

2.1   Data Summary

Primary Research Data Generated by the Project.   As envisaged by the **Serums** Description of Work, the primary data that will be generated by the project will be in the form of the figures and graphs comparing security and privacy of data under existing solutions with the data as used by the**Serums** technologies, based on the chosen metrics outlined in D7.1, and other results that are used to support the research publications that will be produced in the course of the project. The research is expected to yield results in terms of improved security and privacy (measured, e.g., with the amount of leaked data, guessability for the authentication schemes and quality of encryption of the data), reduced system vulnerability, increased trust in the system, increased patient safety etc. Full consideration will be given to allowing proper statistical analysis of the results, using means, standard deviations, regression tests etc. The results, in terms of the chosen metrics, will be stored in textual files and used to generate graphs for research publications. These files will be, for each publications, uploaded to Zenodo (see Section 2.3)

Data Collected by the Project   The output results of the project will be derived from applying the **Serums** technologies on a patient medical data. A part of the results, mostly related to baseline metrics and evaluation of the patient trust in the system, will be derived from questionnaries given to patients and doctors.

5

2.2 Confidentiality and Data Protection Concerns

For most of the project, the research results will be obtained using *synthetic* fabricated data. This data will be generated using the structure of the patient records derived from the real data and will be, from the point of experiments and overall evaluation of the technologies, indistinguishable from the real patient data. On the other hand, it will not be possible in any way to derive the real patient data from the fabricated data. This will allow us to avoid issues of confidentiality of the data for the most part and, therefore, not to introduce any security concerns. The questionnaries will require approval of the relevant bodies (such as National Health Service in the UK) and will not be available for public.

2.3 Data Storage, Access and Curation

The research output data and fabricated data will be, where feasible, be available through Zenodo[2]. This repository has been funded by the European Union with the specific objective of storing the research data that will be generated by Framework 7 and Horizon 2020 research projects. Using this repository has a number of advantages over the self-hosting or use of an institutional repository that we had originally envisaged.

- It provides a more visible and more centralised location for accessing research data, covering a number of EU projects and other sources of research data, rather than a single project and institution. This improves dissemination of the data.

- Data in the repository can be accessed free of charge. A large number of standard licences are supported.

- Data in the repository can easily be searched, mined and otherwise exploited.

- The portal automatically generates a DOI and metadata for each dataset that is submitted. The DOI provides a location-independent reference point that is intended to ensure long-term accessibility of the data. This can be referred to directly from research publications.

- The portal automatically links data to funded EU projects.

- Because it is dedicated to this purpose, the repository provides a long-term storage solution, ensuring the longevity and relevance of the research data. In particular, the data that is generated is guaranteed to be stored and curated long after the project has finished. Without continual intervention, this would difficult to achieve using local and/or institutional repositories.

---

[2] www.zenodo.org

The metadata that is recorded for each dataset will include the date of submission, the owner of the data, a description of the data content, and a link to the **Serums** project.

Preferred Data Formats. To simplify processing and avoid possible problems with transcribing data between evolving data formats, data will generally be stored as ASCII text. A description of the data will be included with each dataset that is stored in the repository.

2.4   Tools for Validating Results

Full information will be provided in the research publications about the tools that have been used to produce the results, including details of operating system versions, libraries and specific software tools, as relevant. The majority of the software tools that will be used by the project will be free, open source software, either produced by third parties or produced by the project and disseminated under open licences. Full care will, however, be taken to avoid releasing proprietory software and to achieve the best possible commercial exploitation for tools that are developed and/or modified in the course of the project. Where tools are restricted, it will generally be possible to obtain these tools under some sufficiently liberal licence agreement to allow full reproduction of the research experiments (perhaps on payment of a fee). We feel that this constitutes the correct balance between obtaining the best possible and most convincing research results, and making those results as open as possible (which taken to an extreme would restrict the data that could be made available/published).