PROJECT DELIVERABLE REPORT

Project Title:

Advanced personalised, multi-scale computer models preventing osteoarthritis

SC1-PM-17-2017 - Personalised computer models and in-silico systems for well-being

Deliverable number	**1.3**
Deliverable title	Data Management Plan Strategy
Submission month of deliverable	6
Issuing partner	UNIC
Contributing partners	ALL
Dissemination Level (PU/PP/RE/CO):	PU
Project coordinator	University of Nicosia (UNIC)
Tel:	**+357 22 841 528**
Fax:	**+357 22 357481**
Email:	**felekkis.k@unic.ac.cy& giannaki.c@unic.ac.cy**
Project web site address	www.OACTIVE.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777159

Revision History

Version	Date	Responsible	Description/Remarks/Reason for changes
1.1	15/4/2018	Kyriakos Felekkis & Christoforos Giannaki (UNIC)	First Draft
1.2	19/4/2018	Anastasia Eleftheriou (CETRI)	Review of first draft
1.3	20/4/2018	Dimitris Tsaopoulos (CERTH)	Review of first draft
1.4	23/4/2018	Kyriakos Felekkis & Christoforos Giannaki (UNIC)	Inclusion of final contribution of the partner's
1.5	24/4/2018	Kyriakos Felekkis & Christoforos Giannaki (UNIC)	Final version for peer review process
1.6	30/08/2018	Kyriakos Felekkis & Christoforos Giannaki (UNIC)	Minor correction

Table of Contents

Abbreviations

EC	European Commission
EU	European Union
OMB	Office of Management and Budget
DMP	Data Management Plan
WP	Work Package

1 Summary

The purpose of this document is to set the Data Management Plan for the OACTIVE project. It contains guidelines that will be used for the development of a Data Management Plan (DMP), which will include an analysis of the main elements of the data management policy that will be used by the OACTIVE consortium with regards to all the data that will be generated by the project. Moreover, the DMP will cover the following aspects:

- Description of the data to be collected / created
- Standards / methodologies for data collection and management
- Ethics and Intellectual Property concerns or restrictions
- Plans for data sharing and access
- Strategy for long-term preservation

The DMP will not be a fixed document, but it will evolve and will gain more precision and substance during project implementation. New versions of the DMP will be created whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors. The first version of the DMP is expected to be delivered in the first months of the project when the first data sets are identified. More detailed versions of the DMP will be delivered at later stages of the project and will be concluded with the final Deliverable D10.4 Data Management Plan.

2 Introduction

2.1 Purpose and Scope

The Data Management Plan (DMP) purpose is to:

• Support the data management life cycle for all data that will be collected, processed or generated by the project.

• Provide an analysis of the main elements of the data management policy, which will be used by the applicants with regard to all the datasets which will be generated by the Project.

• Provide detail and guarantee about the preservation of the data collected during the Project, as well as any results derived from the associated research.

• Provide detail on how the OACTIVE consortium plans to address the ethical issues (if any) related to data, which will be collected during the Project timeframe.

• Create a document, which explains the management of data collected during the Project.

The DMP is not a fixed document, but it will evolve during the lifetime of the Project.

2.2 Intended audience

The DMP will be oriented to:

- OACTIVE project's participant organisations
- Local Ethics Committee
- Partners' personnel and all stakeholders interested in the Project
- European Commission

2.3 Overview of the DMP

The DMP contains details on:

- Brief description of data types which will be collected during the OACTIVE project, explaining the procedures used to collect or create them.
- Copyright and IPR issues.
- Ethical issues related to data storage, persons authorised to see/use them and how long they are kept; managing ethical concerns that include the anonymization of data; procedures used to obtain the consent requested to allow data sharing and reuse.

3 Data Sets Creation for OACTIVE

3.1 Defining research data

One definition of research data is: "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings." [1]. Research data covers a broad range of types of information, and digital data can be structured and stored in a variety of file formats. Note that properly managing data (and records) does not necessarily equate to sharing or publishing that data.

Some examples of research data include:

- Documents (text, Word), spreadsheets
- Laboratory notebooks, field notebooks, diaries
- Questionnaires, transcripts, codebooks
- Audiotapes, videotapes
- Photographs, films
- Protein or genetic sequences
- Spectra
- Test responses
- Slides, artefacts, specimens, samples
- Collection of digital objects acquired and generated during the process of research
- Database contents (video, audio, text, images)
- Models, algorithms, scripts
- Contents of an application (input, output, logfiles for analysis software, simulation software, schemas)
- Methodologies and workflows
- Standard operating procedures and protocols

In addition to the other records to manage, some kinds of data may not be sharable due to the nature of the records themselves, or to ethical and privacy concerns. As defined by the OMB [1], this refers to:

- Preliminary analyses
- Drafts of scientific papers
- Plans for future research
- Peer reviews
- Communications with colleagues

Research data also do not include:

- Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published or similar information, which is protected under law.
- Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.

The following research records may also be important to manage during and beyond the life of a project:

- Correspondence (electronic mail and paper-based correspondence)
- Project files
- Grant applications
- Ethics applications
- Technical reports
- Research reports
- Signed consent forms

3.2 Data sets

The specific data sets for the OACTIVE project need to be identified and described with the contribution of all project partners. A short description of the data, which will be generated in the research project (e.g. samples, physical collections, software, curriculum materials, and other materials to be produced during the course of the project) must be provided. Additionally, an estimation of the amount of data and content of the data (if possible) must be included.

For this reason, the tables with the following data will be filled by the task leaders in order to collect information regarding data sets according to the following template.

TASK	
Data Set Reference & Name	
Data Set Description	Description, Source of data, creation of data,
Standards	word, excel, design etc.
Metadata	Data characteristics
Data Sharing	Data derives from… , Data shared with..., Use of data by…
Archiving and preservation (including storage & backup)	Storage and backups of the relevant materials … first level of storage and backup. e.g. Google Drive folder - second level of storage

All the partners will be asked to provide information regarding the data that will derive from the Work Packages and the Tasks they are leading. For the data sets that will be identified, all the partners will need to provide adequate information regarding the following issues:

Are you generating the data or sourcing it from somewhere else under certain terms and conditions?
Is the data digital or non-digital, or both?
How will the data be created or collected? What instruments or tools will be used to produce the data?
What transformations will the data undergo? What software or file formats will you use as you work with the data?
Will the data be updated or become redundant as you make revisions and produce subsequent versions?
Is the data sensitive or confidential?
Is there ethics approval, or is ethics approval required?

From the information that will be gathered the roles of the partners and the use of the data will be identified. As a result, for each type of research data, it will be defined who will be providing the data and who will be using/analysing the data.

Additionally, the file formats that will be used are an important issue. The formats that will be used should be the best for long-term preservation and continued access of data. Formats most likely to be accessible in the future are:

- Non-proprietary and not tied to a specific piece of software
- Open, documented standard
- Common, used by the research community
- Standard representation (ASCII, Unicode)
- Unencrypted
- Uncompressed

3.3 Descriptive information and Metadata

The DMP defines what documentation and metadata will accompany the data. Metadata is structured information describing the characteristics of a resource; for example, the dates associated with a dataset or the title and author of a book. Metadata supports discovery, re-use and long-term preservation of resources. Metadata needs to vary across scientific fields, but typically cover the following:

- General descriptive and access of metadata

- Data characteristics

- Archive terms and access policies

A metadata record consists of a set of predefined elements that define specific attributes of a resource. Each element can have one or more values; for example, a dataset may have multiple creators. Documenting data enables other researchers to discover your data. Metadata about the nature of the files is also critical to the proper management of digital resources over time.

All the partners will agree on specific issues regarding for example:

- The way that the data will be organised or formatted so that everyone working on it now and in the future knows the origins of the data.

- The way that the each file will be named (File Naming Conventions). The use of the following format is proposed for each file/document: "Date (yyyymmdd)_project_company_filename_author_version". For example, the file containing the minutes from the kick-off meeting will be called: '2017XXYY_OACTIVE_Minutes of Kick Off Meeting_ZZZ_Final'.

- Providing adequate metadata within the dataset (e.g. field labels or column headings) in order to be easy to interpret the data. Other examples of information that the data need to contain include:

  - Reference period
  - Project funding information: European Union logo and information about Grant Agreement and the action/program that funds the project
  - Release policy including dissemination rules and purposes
  - Information about data collection (source, frequency and adjustments)
  - Keywords (Keywords or phrases describing the subject or content of the data)
  - Geographic coverage of the dataset (if applicable)
  - File formats
  - Comments

- Ways to identify different versions. It is proposed in each data set to include a versioning table, additionally to use the prefix ".v1" in each file/document name relevant to the versioning table. For versioning the rule that will be followed will be the use of a sequentially numbered system: v1, v2, v3, etc and "Final" for the final version. If changes need to be done in the final version then the name of the document will change including the relevant sequential version number, ensuring that the document with the "Final" prefix is indeed the final one.

At a minimum, metadata records should be kept in a fielded form, such as a spreadsheet, CSV file, or tab-delimited file. Auxiliary information necessary to interpret the metadata - such as explanations of codes, abbreviations, or algorithms used - should be included as accompanying documentation.

3.4 Ownership (IPR)

In the DMP issues regarding copyright and Intellectual Property Rights of the data are included. This issues are set in the Consortium Agreement and the Grant Agreement of the OACTIVE project regarding all the results of the project. Thus, the DMP follows the Consortium Agreement and the Grant Agreement that is signed by all project partners regarding Ownership issues.

Materials generated under the OACTIVE Project will be disseminated in accordance with the Consortium Agreement. Those that use the data (as opposed to any resulting manuscripts) shall cite and annotate it as follows:

The data were created by the OACTIVE project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 777159. For reuse of this data, please, contact the OACTIVE Consortium. Include your proposed use of the data to assist us in determining your eligibility and to help us navigate possible conflicts between research projects. We will provide you with a short data sharing agreement for you and your authorised institutional official to sign prior to your receiving of the data.

This information must also be described in the metadata

3.5 Storage and access

To ensure the safety of the data, the involved participants will use their available local file servers to periodically create backups of the relevant materials.

Additionally, all other relevant documentation created during the project such as deliverables will be self - archived and preserved in OACTIVE Google Drive that has been created for the purposes of the project. It allows users to store files in the cloud, share files, and edit documents, spreadsheets, and presentations with collaborators. The OACTIVE Google Drive Folder will be accessible by all of the partners of the OACTIVE consortium.

All of the research data and material will be in place for at least 2 years after the end of the project prescribed by the European Commission.

The Coordinator (UNIC) of the OACTIVE project along with the Dissemination & Exploitation Manager (CETRI) will be in charge for data management and all the relevant issues.

3.6 Data Security

The OACTIVE 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;

• Limit the use of USB flash drives, with a clear commitment not to store any personal data on such sticks;

• 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.

4 Ethical and Legal issues

4.1 Ethical Issues

The OACTIVE partners are to comply with the ethical principles as set out in Article 34 of the Grant Agreement, which, among other, states that all activities must be carried out in compliance with:

(a) Ethical principles (including the highest standards of research integrity — as set out, for instance, in the European Code of Conduct for Research Integrity — and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct)

(b) Applicable international, EU and national law

4.2 Confidentiality

All OACTIVE partners must keep any data, documents or other material confidential during the implementation for the project and for four years after end of the project in accordance with Article 36 of the Grant Agreement. Further detail on confidentiality can be found in Article 36 of the Grant Agreement.

5 Conclusions

The document presented the Data Management and Open Access strategy for the OACTIVE project. The Initial OACTIVE Data Management Plan will be developed after the identification of the initial Data sets by all partners. The DMP will be revised and updated during the entire duration of the project. The DMP will be updated at least by the mid-term and final review to fine-tune it to the data generated and the uses identified by the consortium since not all data or potential uses are clear from the start. New versions of the DMP will be created whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors.

6 References

1. Office of Management and Budget (OMB), Uniform Administrative Requirements for Grants and Agreements With Institutions of Higher Education, Hospitals, and Other Non-Profit Organisations, CIRCULAR A-110 REVISED 11/19/93 As Further Amended 9/30/99**,** https://www.whitehouse.gov/omb/circulars_a110#36

7 Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777159.

Disclaimer