Grant Agreement number: **813986**

Acronym: **Syn2Psy**

Title: **Synaptic Dysfunction in Neuropsychiatric Disorders**

Data Management Plan

Initial Version

WP6 – D6.9 – D43

Deliverable date: 11-12-2019

Contents

1. **Document information and Summary** ----------3
2. **Data collection and documentation** ----------4
   - 2.1. Description of the data collected, observed, generated or reused
   - 2.2. Documentation and metadata provided
3. **Ethics and security issues** ----------7
   - 3.1. Handling of ethical issues
   - 3.2. Management of data access and security
4. **Data storage and preservation** ----------9
   - 4.1. Data storage and back-up during the research
   - 4.2. Data preservation plan
5. **Public data sharing** ----------10
   - 5.1. Providing Open Access to data
   - 5.2. Data sharing constraints
6. **Allocation of resources** ----------11
7. **Annex1 (SOP for ESR3)**----------12

1. Document information and Summary

Deliverable Number:	D6.9
Deliverable Title:	Data Management Plan
Work Package:	WP6 - Management
Lead Beneficiary:	CNC
Type:	ORDP: Open Research Data Pilot
Dissemination Level:	Confidential, only for members of the consortium
Due Date:	31-08-2019
Actual Deliverable Date:	11-12-2019
Author	Margarida Figueiredo
Reviewed and Authorized	Ana Luísa Carvalho (coordinator)

Summary

This document describes the initial data management plan (DMP) for the Syn2Psy ITN. The DMP was prepared taking into account the EC Open Research Data Pilot. The data generated during the Syn2Psy project is intended to follow the FAIR principles for data storage and sharing (findable, accessible, interoperable and reusable).

Syn2Psy ITN aims at training a new generation of researchers to unravel the cellular and molecular basis underlying neuropsychiatric disorders (like autism, schizophrenia, intellectual disability) with the goal of contributing to the development of better therapies. A strong focus of this network is in describing the mechanisms by which mutations on synaptic protein-encoding genes cause defects in neuronal networks and in behaviour, since recent studies describe that synaptic pathology is observed in many neuropsychiatric disorders (altered synapse morphology and density, aberrant synapse signalling and synaptic plasticity). The Syn2Psy ITN combines human geneticists with laboratories developing new animal models and with a pharmaceutical company and research groups testing new hypotheses and compounds. The main goals are to: i) generate novel models of disease, including human-derived neurons and brain organoids; ii) test the promising focal point of excitatory/inhibitory imbalance in neuropsychiatric disorders; and iii) explore disease modulatory elements (both positive and negative).

The Syn2Psy research carried out by the ESRs will generate data from studies on: novel animal models created to mimic specific human mutations, human-derived neurons, brain organoids and IPSCs derived from patients as models of neurodevelopmental disorders. The data will include videos of behavioural studies, super-resolution microscopy images and synaptic and mesoscopic imaging, fiber photometry data, optogenetic data and electrophysiology data, among others.

This initial DMP is a living document and its contents will be updated to the end of the project.

2. Data collection and documentation

2.1 Description of the data collected, observed, generated or reused

General data collected and generated:

- Animal data from mice (*Mus musculus*) and rats Wistar Han (*Ratur norvegicus*), including new animal strains (100-150 rats/mice per year for each organization involved).
- Data from Human cells/tissues: iPSCs; dental stem cells extracted from donated teeth from study participants (patients with autism and healthy individuals).
- Viral vectors (Adeno-Associated Virus, lentivirus, etc) and plasmid vectors.

Administrative data (provided as ethics deliverables - due date, February 2020):

- Ethics approvals and certifications for animal experimentation: institution licences (ethics approval for the animal house including authorizations from breeders, suppliers, users), project licences (ethics approval for each project); personal licences (ethics certifications for ESRs - FELASA B minimum).
- Use of Human cells / tissues in research: ethics approvals for the use of cells from patients and informed consent forms; material transfer agreements - MTA (for cells obtained from other laboratories/institutions); details on cell/tissue types and details on biobanks and commercial providers.
- Documentation on health and safety procedures from local/national guidelines/legislation, GMO authorizations, BSL2 measures (security level 2 for work with virus).

Animal husbandry and experimental design data:

- Dark/light cycle, husbandry statistics, animal provenance, mutants.
- Blinding, attrition, randomization, allocation concealment, power analysis, termination criteria, genotyping after animal delivery, genotyping before and after experiment, etc.
- All animal research data will be managed using electronic spreadsheets, animal facility management software **PyRAT®** , **Anibio®, A-Tune** and **Tick@Lab**, and animal behaviour detection and analyses software **Ethovision ®**, **Observer ®**, **Freezeframe**, and Macros designed using **Matlab®**.

Animal physiology and behavioural data:

Behavioural data from living animals and data from *ex vivo* analyses (e.g. brain tissues) will be collected according to standard procedures, and all biological samples will be prepared according to standard published protocols that are well recognized in the field. Data from video tracking of animals' behaviour are expected to reach a total size between 3 TB and 5 TB over the course of each project. The remaining data, together with other derived data (measurements and quantifications), are not expected to exceed 2TB, per project.

Data and Human cells from patients:

For ESR3 project (involving the extraction of dental stem cells from teeth donated by the study participants) we will collect the following data from patients: demographic Data - age, gender; Clinical Data – Diagnosis and Clinical History. The experimental data will be generated from the cell cultures established from the donated tissues (teeth) and from material extracted from them, such as mRNA. The data is expected to have a size < 1 TB. Our data might be useful for other researchers working in similar areas who wish to perform meta-analyses or for comparison of results. Publicly available databases will be re-used for data comparison, such as GTex database and the Allen brain atlas.

Specific types of data generated:

(a) Video files monitoring animal's behaviour during specific tasks
(b) Quantitative analysis data of genes or proteins from brain tissue
(c) Quantitative analysis data of hormones from blood samples
(d) Quantitative analysis data on brain metabolites with magnetic resonance spectroscopy
(e) Polysomnographic Electroencephalogram/ Electromyography (EEG/EMG) recordings
(f) Field and whole-cell electrophysiological data on neuronal cell activity *ex vivo* and *in vivo*.
(g) RNA expression data, including data from ribosome profiling analysis (RNAseq), Drop-seq and qRT-PCR
(h) Protein expression data
(i) Flow cytometry data
(j) Immunocytochemistry data
(k) Whole-cell patch clamp data/field recordings
(l) Two-photon microscopy data
(m) Data obtained by super-resolution imaging methods (PALM, 2-color 3D Storm, UPaint)
(n) Data from mesoscopic optical voltage imaging
(o) Data from real-time quantum dot tracking of single receptors
(p) Data from monosynaptic retrograde tracing
(q) Data from magnetic resonance diffusion tensor imaging (MRI-DTI)
(r) Demographic Data - age, gender and Clinical Data - Diagnosis, Clinical History (ESR3 – using autism patients dental stem cells)
(s) Neuronal morphometric analyses - Sholl
(t) Neuronal activity measurements (GCaMP)
(u) Ultrasonic Vocalizations (USVs) recordings

Syn2Psy data files:

When possible data will be saved as commonly used file types, e.g. .**tiff**, **.jpeg, .xlxs, .png, .txt** and **.xlm**. For video tracking, files will be saved as .**avi**, **.mp4** and **.raw**. Below are examples of file types that might be generated from specific software and techniques:

.lsm and **.czi** files(**ZEISS** microscope format) - confocal data;

.lif, **.lifext** (**Leica SP8** microscope format) - confocal data;

.tiff files (**ImageJ** and **Prairie View**) - *in vivo* 2-photon imaging, macros, analysis & processing of images; **.vsi** - scanned slides;

.sds - gene expression;

.prm - chromatography primary files for HPLC;

.pvfs (**Sirenia**® software) - *EEG*/EMG;

.dat, **.hdf5** (**Stimfit** software) and **.pClamp** - electrophysiology;

.png, .txt, .xlm (**Neuroexplorer** and **Openphys** software) - *in vivo* electrophysiology recordings;

.pxp or **.ibw** (**WaveMetrics Igor Pro**® Software) - to analyze electrophysiology data or exported as **.csv;**

.xlxs, **.png** or .**xlm** (**GraphPad-Prism** Software) - to make graphics of electrophysiology data;

.ab1, **.seq,**.**fastq** - sequencing data;

.gel (**Typhoon** software) - laser scanner data;

.ims (**IMARIS** software files) - Sholl and spine analyses;

.m (acquired data from **Matlab** custom software scripts) – animal behaviour;

.imu - accelerometer data;

.gpio and **.isxd** (**Inscopix** software) - DAQ and for ancillary data respectively;

.wav and .**aiff** (**Audacity** and **Raven Lite** software) - USVs;

.xlsx - measurements of behavioural parameters (for long-term preservation: **.csv**);

.ffii (**Freezeframe** Software) - behaviour data.

2.2. Documentation and metadata provided

Nomenclature:

Datasets will be stored with criteria that will ensure accessibility to all current and future members of the research group and that will facilitate future consultation and reuse of the data. Files will be named according to a pre-agreed convention that will include identification of Syn2Psy, researcher name, date and the type of data, e.g. **Syn2Psy_ResearcherName_date_DataType**'.

We will provide keywords in the published scientific articles. We will follow the Human Genome Organisation naming conventions, defined by the HUGO Gene Nomenclature Committee (HGNC).

Unique identifiers:

Animals used in each dataset will be identified by the **unique code** assigned by the animal facility software **PyRAT®, Anibio®, A-Tune** and **Tick@Lab** or blindly by the animal facility personnel**,** allowing access to genetic background and pedigree.

For iPSC cells derived from patients by collaborators, the collaborators coding system for the cells will be used (designated by clone number and cell type [disomic or trisomic]). All other identifier information pertaining to patients will be kept blinded from the researcher.

For ESR3 project (involving the extraction of dental stem cells from donated teeth of the study participants): each patient sample is assigned a Project ID number unique to each patient, which will be provided in a pseudonymized form along with relevant information (age, gender, diagnosis).

Syn2Psy Publications, underlying data, and corresponding metadata/bibliographic metadata will be assigned a **DOI (Digital Object Identifier)**. All Syn2Psy researchers will have an **ORCID** number (non-proprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors).

Metadata

Each dataset will be accompanied by a **README.txt** file which will describe the directory hierarchy, the purpose, origin, description, time reference, creator, access conditions and terms of use of a data collection and providing details on the experimental procedures.

The final datasets, as deposited in the chosen data repositories, will be accompanied by the relevant metadata documentation that will be compliant with **DataCite Schema** standards. This will ensure that the produced data can be analysed together with other data types during the project and following project completion.

Bibliographic metadata, which identify publications deposited in open access repositories, will include: [EU_Horizon2020_MSCAITN_Syn2Psy_GA813986_PublicationDate_LenghtEmbargoPeriod_DOI].

Documentation procedures:

Along with the stored data, we will annotate what has been done in detail: the conditions of the analytical procedures; sample type, size, date of analysis, fixed and variable modification done during the sample preparation, method used for sample preparation, units of measurements, version of software used, the date on which each modifications were made, the type of modification or data processing, if any.

3. Ethics and security issues

3.1. Handling of ethical issues

All research work proposed in Syn2Psy will be conducted in accordance with the relevant national rules of the country where the research will be carried out. Whenever necessary, authorization will be requested by the competent research ethics committee or the medical ethical committee of the country where the work is conducted. The proposed work has implications on two different levels: the use of (i) laboratory animals and (ii) human-derived samples, from healthy volunteers and patients with *Shank3* mutations.

Animal Experimentation:

All Syn2Psy research groups have adhered to the European Commission guidelines for the care and use of laboratory animals, which have been acquired and transformed into operative rules by the national (Ministry of Health or Ministry of Agriculture) and local institutions (Health Services, Universities and their operative branches, the Departments and Institutes), and monitored by local Ethical Committees.

Research involving animals - small rodents, in the case of the present project - will be performed in compliance with Directive 2010/63/EU (revising Directive 86/609/EEC), which is transposed into the national law of all partners involved.

The **Syn2Psy** project will comply with the *Amsterdam protocol on animal protection and welfare*, which is characterized by the need to **Replace** animals whenever possible, to **Reduce** the number of animals used in experimental research, and to **Refine** procedures to minimize suffering (3Rs approach). The following principles will be implemented: **Reduction:** we will make all efforts to *reduce* the number of animals used in experiments to the minimum that is required for obtaining meaningful results. **Replacement:** A major effort will be made to keep animal experimentation to the minimum required and to *replace* the use of live animals by non-animal alternatives. **Refinement:** Mice (wild type and genetically modified) will be housed in certified animal facilities at all participating institutions and all effort will be made to minimise animal suffering and distress during handling and experimentation.

In order to ensure the best practice in experimental research with animals, all **ESRs** will be required to **successfully conclude a FELASA course - Category B, or equivalent** (for persons carrying out animal experiments).

Human cells from patients

For ESR3 project (involving the extraction of dental stem cells from teeth donated by the study participants): Only the clinical staff has direct access to patients and all medical information. The clinicians explain the overall goals of the project, recruit participants, store informed consents, extract biological samples and perform pseudonymization. The research team is given accessed to limited data from the participants, which is mostly restricted to age, gender, clinical condition and relevant biological samples. The pseudonymized personal data of the patients (e.g demographics and clinical data) is kept alongside participants' IDs and other relevant information related to the biological sample (e.g. passage number, cell types, etc.). Only the healthcare team at the Pediatric Hospital will have access to the key that can identify the participants. No identifiable personal data will be made available to the research team or for publication. The biobank responsible for storing the samples manages its own license for use of samples for research and is under the purview of the Portuguese Ministry of Health. The informed consent given to participants clearly states that generated data can be shared within the research community and states the above-mentioned safeguards regarding personal data protection. Regarding long-term preservation of data and biological samples, the Portuguese Legislation indicates that research study participants cannot require erasure or destruction of samples.

3.2. Management of data access and security

All computers in the laboratory are only accessible through a session opened by users. User accounts (user_ID) and passwords are managed by the Institution's IT Departments. Access to the various servers in the laboratories is assigned to users according to their needs and controlled by the PI.

Human cells from patients (ESR3)

For ESR3 project (involving the extraction of dental stem cells from teeth donated by the study participants): the best security practices will be followed in order to ensure the **confidentiality**, **authenticity** and **integrity** of the information. The participants' pseudonymised personal data (e.g. demographics and clinical data) is kept in a password-protected platform- *S.Clinico* - that is maintained by the Portuguese Government through the Ministry of Health, and is accessible only by clinicians at the Paediatric Hospital. Project-specific information is transferred to *CliniData*, a platform associated with the Paediatric Hospital Biobank, where the participant ID/codes is deposited. Researcher staff only has access to pseudonymised data in the *CliniData* platform.

The research team does not have access to the key that can identify the participants. The healthcare team (primary caregivers) at the Pediatric Hospital retains all patient data and clinical history and deliver the relevant information (age, gender, diagnosis) to the research team in an associated, pseudonymized form, alongside an Project-specific ID number unique to each patient.

Research data will be archived in a durable form for a period of 10 years with the research team in password-protected offline storage. The team has implemented a Standard Operating Procedure (SOP) (in Annex 1) to cover handling of samples and assure the maintenance of patient confidentiality. This outlines the controls and checkpoints that are in place when processing different types of confidential information and biological samples.

Syn2Psy Intellectual Property (IPR)

The rights of beneficiaries and partner organisations to use the knowledge produced within Syn2Psy are subjected to IPR clauses in the **consortium agreement** that was signed by all parties. The Syn2Psy consortium will examine and operate any potential IPR issues as they happen.

4. Data storage and preservation

4.1. Data storage and back-up during the research

Animal husbandry data from **PyRAT®**, **A-TUNE**, **Anibio®** and **Tick@lab** software will be stored at the animal husbandry units and exported and stored at the institutional servers.

Data is collected in original notebooks and hard- copies of quantification will be stored in the PI's laboratory or office. Additional electronic data will be stored on the individual experimenter's computer and backed up on external disks / Lab Server only accessible from the research group. Additionally, the laboratory will make use of the PI's lab server space at institution's storage facility for a second repository of data storage (when available). Data will be preserved for 10 years on the Institution's servers.

4.2. Depositing data in repositories

For long-term data preservation, Syn2Psy research results in the form of peer reviewed final manuscripts and in the form of underlying data (data required to validate the results described in the scientific publication and corresponding metadata) will be immediately deposited in a digital **data repository** conforming with the FAIR data principles (e.g. Zenodo), at the latest on the publication day. Bibliographic metadata, which identifies the publications, will also be deposited. Manuscripts ready for peer-review (preprint) might in some instances be released as open-access preprint on **BioRxiv**.

Datasets will be given a Digital Object Identifier -**DOI**- and associated metadata. The **DOI** corresponding to the datasets in the repository will be included in the article's reference list, allowing identification and access of any dataset in a publication.

DOIs will also be linked to appropriate records in the Institution's publication repository, to enhance the dataset's visibility. Metadata about datasets will be publicly searchable and discoverable and will indicate how and on what terms the datasets can be accessed. Iinformation about datasets will be displayed on the laboratorys' webpage, on researcher profile pages, as well as on researchers' **ORCID_iDs**, which will increase the visibility of the datasets.

After depositing the data, open access to the publications will be ensured.

Unpublished high-quality final data generated during this project will be saved for reuse in the research groups' future projects. We will make some data available for use by researchers in future collaborations if our lab no longer uses it. Bad quality data will be permanently discarded at the end of the project. Where possible, we will store files in open archival formats e.g. **.docx** files converted to **.pdf-A** or simple **.txt** files encoded in **UTF-8**, **.xlsx** files converted to **.csv**, **.LSM** and **ImageJ** files to **.tiff**. In case this is not possible, we will include information on the software used and its version number.

Coding data and custom software will be shared by deposition on appropriate online repositories (e.g. direct link to lab server, Bitbucket, GitHub, and ModelDB), and custom design files will also be made available (e.g. www.thingiverse.com and GitHub).

5. Public data sharing

5.1. Providing Open Access to data

Syn2Psy will, as far as possible, enable the generated data to be openly accessible, free of charge, allowing the users to mine, exploit, reproduce and disseminate it (unless specified in this document or in future Syn2Psy data management plans).

Syn2Psy will provide **open access** to published data, and corresponding bibliographic metadata, via self-archiving (in **Zenodo** and Institution's data repositories) no longer than 6 months after the publisher's publication date (green open access) or immediately upon publication (gold open access – via open access journals or by paying additional fees in hybrid journals). Syn2Psy will aim at establishing CC BY license with the publishers regarding the researchers/authors copyrights over their work. The EC Open Access Publishing Agreement: V1.0 – 20.03.2017 might be used, when necessary, to ensure that the publisher's embargo period for public dissemination is not longer than 6 months.

Underlying data (needed to validate results presented in scientific publications), and corresponding metadata, will also be openly accessible, through deposition in **Zenodo**.

5.2. Data sharing constraints

Syn2Psy will, in general, have no data sharing constraints, except when using data from patients due to the need to protect personal data and assure their confidentiality and security (see below). All chosen digital repositories are/will be conformed to the FAIR Data Principles (e.g. **Zenodo**).

Data from patients:

For ESR3 project (involving the collection of human dental stem cells from autist patients), to assure participant confidentiality, the following data will not be publicly available: Demographic Data, Clinical History. **Neuromorpho.org** will be the data repository used for morphometric data. Since the data will be stored in a public repository and used for scientific purposes, the identity of those accessing the data will, therefore, not be ascertained. Our data will be uploaded without any identifiers, tags or metadata, regarding participant identity. The available data produced in the project will be interoperable, allowing data exchange and re-use between researchers, institutions, organizations and countries. The data will be licensed through creative commons attribution and non-commercial licenses. Data will be available upon publication through scientific articles, which will include the appropriate link to find the relevant datasets. The shared data will be re-usable, since it will be deposited in public repositories and, upon request, given to the leaders of the project. The data will remain re-usable during the lifetime of the public repositories and offline back-ups will be kept for a period of 10 years. Data quality assurance will be described in the methods section of the associated scientific articles.

6. Allocation of resources

The costs of making the Syn2Psy data FAIR are related to the time spend in these activities and are estimated to be about 1 week/year for each partner. There is a continuous need to update and maintain informatics hardware associated with the maintenance of data storage. Those costs are funded by project funds.
Every Syn2Psy researcher is responsible for the management of the data that he/she will collect/generate. The project manager will overview the implementation of this data management plan.

For the costs related to preservation of digital data, Syn2Psy will use the free data repository **Zenodo**.

Preservation of biological samples will be the responsibility of the partners that generated them. Important samples can be transferred to third-party repositories that do not charge for archiving (e.g. Adgene).

Annex 1

Standard Operating Procedure (SOP) for ESR3 Project (involving the collection of human dental stem cells from patients with autism)

Below we describe the procedures, since sample collection until data preservation:

1) Participant personal and clinical data is collected by clinicians and stored in S.Clinico Platform.

2) Informed consent form is given and signed during medical consultation with the medical staff.

3) In case of deciduous teeth, parents take a tooth kit home, which includes a tube with medium to preserve the dental cells and instructions to have the sample kit delivered to the Paediatrics Hospital.

4) In case of third molars, a surgical excision is scheduled.

5) Both the tube and consent form are coded by attaching a tag with an ID code.

6) Once the tooth is shed or collected, it is placed inside the coded tube and brought to the Paediatric Hospital by the parents or clinical staff.

7) The coded tube is then picked up by the research team and sent to the lab at CNC, University of Coimbra.

8) The sample is cultured and expanded by cell culture procedures.

9) Frozen sample is then sent to the Biobank at the Paediatric Hospital for storage, where the identification of participants is stored in their platform CliniData.

10) Pseudonymized data generated in the lab will be stored in password-protected drives.

