IMMERSE

Implementing Mobile MEntal health Recording Strategy for Europe
H2020 - 945263

D6.2	**POPD - Requirement No. 4**
Dissemination level[1]	CO
Contractual date of delivery	30.6.2021
Actual date of delivery	30.6.2021
Type[2]	R
Work package	6
Work package lead	Ine Van Hoyweghen

Author list

Organisation	**Name**	**Contact information**
1 - KU Leuven	Luca Marelli	luca.marelli@kuleuven.be
9 - TMF	Irene Schlünder	Irene.Schluender@tmf-ev.de
1 - KU Leuven	Ine Van Hoyweghen	ine.vanhoyweghen@kuleuven.be
6 - UHEI	Thomas Gastland	Thomas.Ganslandt@medma.uni-heidelberg.de
2 - CIMH	Ulrich Reininghaus	Ulrich.Reininghaus@zi-mannheim.de
2 - CIMH	Lena Wieland	Lena.Wieland@zi-mannheim.de
2 - CIMH	Anita Schick	Anita.Schick@zi-mannheim.de
3 - UEDIN	Matthias Schwannauer	M.Schwannauer@ed.ac.uk
3 - UEDIN	Tariq Elahi	t.elahi@ed.ac.uk
5 - UPJS	Iveta Nagyova	iveta.nagyova@upjs.sk
8 - Movisens	Simon Krause	simon.krause@movisens.com

[1] **Please choose the appropriate reference:**
PU = Public, fully open, e.g. web;
CO = Confidential, restricted under conditions set out in Model Grant Agreement;
CI = Classified, information as referred to in Commission Decision 2001/844/EC.

[2] **Use one of the following codes:**
R: Document, report (excluding the periodic and final reports)
DEM: Demonstrator, pilot, prototype, plan designs
DEC: Websites, patents filing, press & media actions, videos, etc.
OTHER: Software, technical diagram, etc

2 - CIMH	Manuel Brenner	Manuel.Brenner@zi-mannheim.de
4 – UK BA	Anton Heretik	anton.heretik2@uniba.sk
4 – UK BA	Michal Hajduúk	michal.hajduk@uniba.sk
3 - UEDIN	Maria Wolters	Maria.Wolters@ed.ac.uk
2 - CIMH	Georgia Koppe	georgia.koppe@zi-mannheim.de

Document history

Version	Date
1	30.6.2021

List of abbreviations

N/A

Table of contents

1. Summary

This deliverable outlines key measures implemented in the IMMERSE project to comply with data protection requirements. It provides a declaration of compliance with national data protection legislation of the countries where the research takes place, as well as confirmation of the appointment of Data Protection Officers (DPOs). The Deliverable also outlines the data minimization strategy that will be adopted in the course of the project, along with the technical and organizational measures that will be implemented to safeguard the rights and freedoms of the data subjects, including description of anonymization/pseudonymization techniques. Finally, it provides confirmation of legal compliance for data transfers to/from the EU to the UK.

2. Deliverable report

2.1 The beneficiary must check if special derogations pertaining to the rights of data subjects or the processing of genetic, biometric and/or health data have been established under the national legislation of the country where the research takes place and submit a declaration of compliance with respective national legal framework(s).

We hereby declare that the intended data capture and data use within IMMERSE is compliant with the national laws of our Consortium Members.

Specifically, for **Germany** there is no derogation for health data or other special law on research involving human participants. The applicable general data protection law (Landesdatenschutzgesetz Baden-Württmberg https://www.landesrecht-bw.de/jportal/?quelle=jlink&query=DSG+BW&psml=bsbawueprod.psml&max=true&aiz=true) and the professional law for clinicians (Berufsordnung der Landesärztekammer Baden-Württembergh https://www.aerztekammer-bw.de/10aerzte/40merkblaetter/20recht/05kammerrecht/bo.pdf) is being observed.

For **Belgium**, the applicable data protection 'framework law' on data protection (Wet betreffende de bescherming van natuurlijke personen met betrekking tot de verwerking van persoonsgegevens 30 Juli 2018, https://www.autoriteprotectiondonnees.be/publications/loi-cadre.pdf) is being observed, notably with respect to relevant provisions contained in Title 4, articles 186 - 208 (TITEL 4. — Verwerking met het oog op archivering in het algemeen belang, wetenschappelijk of historisch onderzoek of statistische doeleinden bedoeld in artikel 89, §§ 2 en 3, van de Verordening).

For **Slovakia**, the applicable data protection legislation (Act no. 18/2018 Collection of Laws of the Slovak Republic on Personal Data Protection, https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2018/18/20180525) is being observed.

For the UK, the applicable data protection legislation is being observed (Data Protection Act 2018).

2.2 The host institution must confirm that it has appointed a Data Protection Officer (DPO) and the contact details of the DPO are made available to all data subjects involved in the research. For host institutions not required to appoint a DPO under the GDPR a detailed data protection policy for the project must be submitted as deliverable.

Partners responsible for collecting personal data in the course of the project, which will act as data controllers upon definition of the data governance framework within the consortium (due for Month 6), have appointed the following Data Protection Officers (DPOs):

Central Institute of Mental Health Mannheim:
Frau Dr. Regina Mathes
Contacts: Tel. +49 6 221 56 - 7036; E-Mail: datenschutzbeauftragter@zi-mannheim.de

KU Leuven:
Dr. Griet Verhenneman
Contact: griet.verhenneman@uzleuven.be

Comenius University Bratislava:
Mr. Ladislav Kovar, MA.
Contact: ladislav.kovar@uniba.sk

UPJS Kosice:
Ms Gabriela Ciberejova.
Contact: Tel.: +421 55 234 1586; E-mail: gabriela.ciberejova@upjs.sk

University of Edinburgh:
Dr. Clare Whittaker
Contacts: Phone +44(0) 131 651 4052; Email: clare.whittaker@ed.ac.uk

In line with GDPR requirements (arts. 38 and 39 GDPR), the DPO will advise data controllers on their legal obligations, will monitor compliance with the GDPR and national legislation, and will act as contact point for data subjects with regard to all issues related to the processing of their personal data and to the exercise of their rights under the GDPR and national legislation. The contact details of the relevant DPO will be provided to research participants in the informed consent template.

2.3 The beneficiary must explain how all of the data they intend to process is relevant and limited to the purposes of the research project (in accordance with the 'data minimisation 'principle). This must be submitted as a deliverable.

Phase I (months 1-18):
In Phase I of the project, a participatory field study will be conducted using a questionnaire (part A) as well as qualitative semi-structured interviews (part B) with service users, clinicians, the service users' support network, health care system administrators and managers. Questionnaire data (part A) will be collected anonymously. The data processed in the study for semi-structured interviews (part B) will be pseudonymised. In Part A and Part B, service users have the opportunity to share

information about their diagnosis, symptoms, and use of technology to manage their mental health. Service users are fully in control of the type, level, and amount of health information they share.

Specifically, Phase I of the study will process the following types of personal data, including special categories of personal data such as health data (art. 9 GDPR):

- Information about the mental health of service users in Parts A and B, if given
- Audio files (if permission to record has been given) of interviews in Part B, transcripts of interviews pre-pseudonymisation.

All data will be collected pursuant to the ethics and regulatory pathways described in the study protocol.

The questionnaire items in Part A and the topics covered in the interview in Part B have all been carefully mapped onto the Non-Adoption, Abandonment, Scale-Up, Spread, Sustainability framework (NASSS, Greenhalgh et al., 2017) to ensure coverage of all seven dimensions of the framework. In addition, views on planned implementation strategies for Phase II will be probed. The questionnaire covers six of the seven NASSS dimensions using brief versions of established instruments, custom items, and a small number of optional free text questions. Implementation strategies will be covered by custom questionnaire items and illustrated during interviews with a small number of vignettes. For both part A and part B, demographic information will be collected in a way that minimizes the risk of being able to identify participants by mapping the demographic data collected onto the service users, clinicians, supporters of service users, and health service administrators of each country and site.

Phase II (months 18-48):
In **Phase II** of the project, a multi-centre, parallel-group cluster randomized controlled trial (cRCT) will be carried out, in which 24 clinical units (as the cluster and unit of randomization) within mental health services at eight sites in four European countries are randomly allocated to one of two conditions: (a) the experimental condition, in which participants receive the intervention in addition to treatment as usual (TAU) or (b) the control condition, in which service users are provided with TAU.
The data processed in Phase II of the project are composed of personal data that provide information on the state and changes in patient's symptoms, context, and their interplay, and which will be collected pursuant to the ethics and regulatory pathways described in the study protocol. Specifically, the study will process the following types of personal data, including special categories of personal data such as health data (art. 9 GDPR):

1. ESM data, provided by service users through the ESM app. These include self-reported momentary mental states and symptoms.
2. eCRF data, collected by the investigators in the context of the trial.
3. Audio files (if permission to record has been given) of interviews, transcripts of interviews pre-pseudonymisation.
4. Mobile sensing data (such as steps, physical activity, app usage, location), captured by the movisenseXS app, which will be used and integrated through an Application Programming Interface (API) to the Research Database.

Within the broader discussions over the definition of a Data Governance Framework and Data Management Plan for IMMERSE (Milestone deadlines at M6), project partners are presently undertaking regular consortium-wide discussions to: (i) identify the *specific* types of data that will be relevant to collect and use in Phase II of the project, according to the research questions being pursueds as well as the state-of-the-art in the field; and (ii) ensure that the data processed within Phase II of the project falls within the remit of the data minimization principle, as enshrined in the GDPR (Art. 5.1.c).

The data minimization principle refers to both the scope and categories of data initially collected. In line with the purpose limitation requirement, this principle is strictly related to the intention of avoiding the buildup of extensive data collections that may lead to risks of social surveillance and control (see e.g. Mantelero 2014; Zarsky 2017), as well as minimizing the possibility of data breaches, whose consequences can be hugely detrimental to data subjects (see, e.g. O'Doherty et al. 2016; Pasquale 2015, 29).
Yet, it is widely acknowledged that, while representing a traditional pillar of data protection legislation, the principle of data minimization maintains an inherent tension with current developments in digital health, notably with respect to the implementation of machine learning approaches (see e.g. Marelli et al. 2020). Epistemically, inasmuch as machine learning algorithms "learn and develop" (Kuner et al. 2017) and largely follow a data-driven (rather than hypothesis-driven) model, it can be difficult to foresee in advance the very purpose of the processing of personal data, or, conversely, to precisely identify at the onset of a research the types of data that will be necessary to use to achieve the envisaged outcome. In the same vein, while the principle of data minimization is geared to reduce the amount of data processed, machine learning analytics work at its best when making combinatorial and repeated use of the data, which in turn hinges on the collection of vast quantities of personal data.

These well-known challenges are central to the research carried out within IMMERSE, which sets out - accordingly - to achieve the right balance between these possibly competing requirements, with the intent to maximize research outcomes in compliance with this fundamental data protection principle.

As a *general approach*, in order to minimize data collected during the study, items will only be included when specifically relevant for the planned analyses. Also with regard to mobile sensing data, the specific data elements to be captured will be determined based on concrete relevance to the research topic in order to ensure data minimization and to ensure, to the best of available knowledge and techniques, that no tracking, surveillance, profiling or scoring of service users can be implemented.

More specifically, IMMERSE will **create and maintain a list of essential data items** that need to be collected in the project, based on current state-of-the art in the field, the research questions the project intends to address, the technical affordances of the movisenseXS app, which will be used to collect mobile sensing data. The list (See **Annex A**), which may undergo revisions as the project progresses, is composed of different types of data, which need to be processed in the project according to the stated rationale.The list will undergo periodic assessment and possibly revisions as the project

progresses. Partners of the different WPs will be involved in such assessment, geared to constantly uphold the principle of data minimization.

2.4 A description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants must be submitted as a deliverable.

The data processing activities carried out within the scope of the present project will be based on the explicit consent of data subjects, as detailed in D6.3, part 2. However, it is widely acknowledged that such an "informational self-determination" or "notice and consent" approach - which is rooted in a distinctively liberal paradigm that conceives of the idea of privacy as an individual right to control the use of one's own data - is not effective enough, *per se*, in providing adequate safeguards to data subjects (Marelli et al, 2020). Accordingly, a number of additional organizational and technical measures need to be implemented to ensure that data processing activities occur within a robust ethical framework safeguarding the rights and interests of data subjects.

Organizational safeguards measures
The following organizational measures will be implemented to safeguards the rights and freedoms of data subjects:

(i) Ethics review: the project will undergo full ethics review by the local competent research ethics committees at KU Leuven, CIMH, UEDIN, UK BA, UPJS, UHEI, UKHD.

(ii) Appointment of a Scientific and Ethics Advisory Board (SEAB). An external independent Scientific and Ethics Advisory Board has been appointed, and is tasked with monitoring the ethical and legal issues involved in the project and how they are handled, including those revolving around the processing of personal (sensitive) data. Details on the composition of the SEAB, the rationale for its appointment, and the tasks it carries out, are provided in D6.5.

(iii) Appointment of **Data Protection Officers (DPOs).** Each partner in charge of data collection and acting as data controller within the project has appointed a DPO, as detailed in point 2 above.

(iv) Implementation of a **data governance framework**: As outlined in the GA, by Month 6 the project will implement a comprehensive data governance framework, to define rules and procedures for each primary and secondary data processing activity carried out with personal (sensitive) data collected in the project. Within such a framework - currently under definition among project partners - a **Data Governance Board (DGB)** will be established. The DGB will be composed of the members of the Steering Committee, and will be in charge of overseeing and approving all data processing operations and data flows within and outside the IMMERSE consortium. It will thus be responsible for exercising data governance throughout the full life cycle of collected personal data, while aiming to valorize personal data processing for research in compliance with ethics and normative requirements.

The Data Governance Board will be tasked with implementing a ***two-layered access model***, according to the following principles and criteria:

(i) For requests of data access by members of the IMMERSE consortium, de-identified data (including appropriate FAIR metadata annotations) will be made available through an **automated data check-out system**. The system will require researchers to pre-register their research hypotheses using the Open Science Framework and submit a request for the data necessary to answer the a priori defined research questions. Upon abstract approval, a dataset containing only variables required for the proposed analysis is then released to the researchers, along with a time- and date-stamped receipt of data access.

(ii) For requests of data access by researchers outside the IMMERSE consortium, or by members of the IMMERSE consortium wishing to use data for secondary research purposes outside the scope of the IMMERSE protocol, the DGB will be tasked with implementing oversight and **controlled access functions**. Controlled access review criteria, still being discussed within the consortium, will likely include (a subset among) the following requirements as pre-condition for data access: (i) Affiliation to a recognized research institution; (ii) qualification to undertake the proposed analyses; (iii) Compatibility of proposed study with IMMERSE objectives; (iv) Compatibility of proposed study with the consent requirements; (v) Required ethical obligations have been met; (vi) Scientific merit of proposed study (clarity, novelty and scientific excellence); (vii) Applicants' privacy and confidentiality policy and security measures are adequate; (viii) Applicants' institution has approved and signed a Data Access Agreement.

Technical measures
Alongside organizational safeguard measures, IMMERSE will implement technical safeguards. In particular, the study will adhere to a strict **pseudonymization** policy, as described in pt. 5 below. It will also implement state-of-the-art **security measures** to prevent unauthorized access to personal data, as described in D6.3.

2.5 Description of the anonymisation/pseudonymisation techniques that will be implemented must be submitted as a deliverable.

Principles
The study will adhere to a strict pseudonymization policy, by strictly implementing a technical and organizational separation of identifying and medical data elements. A central ID- and Consent Management platform enforces such a strict separation of identifying and medical data elements, with pseudonymized datasets in the research components of the platform (eCRF and DDMH Research Database). The research database will thus contain only pseudonymized data. Exception to this strict pseudonymization policy is represented by service users and treating physicians, which need to be able to see identities within the treatment context.

Importantly, specific attention will be paid to highly identifiable types of data (such as free-text data, or data related to geolocalization), that may lead to re-identification of data subjects or profiling. As a general principle, the consortium will not include highly identifiable data points in the research component of the platform. The exceptions are properly pseudonymised free text and interview data, where all person names, place names, and events that might lead to the identification of participants are substituted with pseudonyms or omitted.

Data provided to external researchers will be anonymized by removal of identifiers or pseudonyms as well as generalization, perturbation or suppression of payload data when required to prevent re-identification. Moreover, periodic assessments will be made to keep track of the data being shared, in order to prevent cross-linking of different datasets that may lead to re-identification of data subjects. For example, previously shared data could prevent the consortium from sharing other data with the same third party, should the combination of the two datasets heighten the likelihood of re-identification of data subjects.

Techniques

We now describe the technical aspects of the techniques the consortium will employ to protect the privacy of data owners, depending on the context and use-case.

Pseudonymization: The names of participants will be replaced by random strings (format and length to be finalized). These strings can not be reversed due to the use of cryptographic hash functions. The mapping of pseudonym to true name will be stored with the eCRF system with access control as described above. IP addresses, and other quasi identifiers, will be pseudonymized in a similar fashion, however, the mappings for real values will not be retained in the system. The names of places and events will be pseudonymised in a way that prevents identification, but preserves information about type and place of events. Years and dates from interviews will be replaced by the nearest decade. A preliminary procedure for this is to be finalised and will be constantly revised during collection and analysis of free text and interview data to cover any gaps.

Inference control

A number of specific controls will be utilized to limit the amount of private information leakage from ongoing research activities.

***Suppression*:** Identifying information in free-text fields will be deleted before the data is copied for research purposes (i.e. not medical purposes). The research database may also suppress data fields such as the location, time, and certain statistical measures depending on the context of the query.

***Aggregation*:** Statistical queries to calculate summary statistics such as the mean, median, and quartiles shall be restricted to requirements of minimum anonymity set sizes. This is to prevent leakage about individuals through aggregated statistics. The anonymity set size will be adjusted to the particular scenario of the use-case and will depend on the statistic being calculated.

Differential privacy: Numerical data, both for direct use or for machine learning purposes, will have noise added either at the source or before storage in the master

database (i.e. original values will not be stored). This noise will be tuned to the utility requirements of the tasks and sensitivity of the statistics used in the query.

2.6 In case personal data are transferred from the EU to a non-EU country or international organisation, confirmation that such transfers are in accordance with Chapter V of the General Data Protection Regulation 2016/679, must be submitted as a deliverable.

2.7 In case personal data are transferred from a non-EU country to the EU (or another third state), confirmation that such transfers comply with the laws of the country in which the data was collected must be submitted as a deliverable.

IMMERSE will transfer data to and from the UK (University of Edinburgh). At the time of writing (June 26), however, it is not yet clear whether the European Commission will adopt an adequacy decision for transfers of personal data to the United Kingdom. Such an adequacy decision is expected to be adopted soon, and will ensure that such data transfers are in compliance with European data protection standards. In case no adequacy decision is reached, the University of Edinburgh has contingency plans in place. Namely, they have drafted templates for data transfers agreements (DTAs) to ensure that data transfers to and from the UK are in compliance with respective data protection legislation. We expect to be able to provide a confirmation on these two requirements within the next month.

3. References

Mantelero, A. 2014. “The Future of Consumer Data Protection in the E.U. Re-Thinking the “Notice and Consent” Paradigm in the New Era of Predictive Analytics.” Computer Law & Security Review: The *International Journal of Technology Law and Practice* 30 (6): 643–660.

Marelli, L., Lievevrouw, E., & Van Hoyweghen, I. (2020). Fit for purpose? The GDPR and the governance of European digital health. *Policy Studies*, 1–21.

O’Doherty, K. C., E. Christofides, J. Yen, H. B. Bentzen, W. Burke, N. Hallowell, B. A. Koenig, and D. J. Willison. 2016. “If You Build It, They Will Come: Unintended Future Uses of Organised Health Data Collections.” *BMC Medical Ethics* 17 (1): 54.

Pasquale, F. 2015. *The Black Box Society. The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Kuner, C., D. J. B. Svantesson, F. H. Cate, O. Lynskey, and C. Millard. 2017. “Machine Learning with Personal Data: Is Data Protection law Smart Enough to Meet the Challenge?” *International Data Privacy Law* 7 (1): 1–2.

Zarsky, T. 2017. “Incompatible: The GDPR in the Age of Big Data.” *Seton Hall Law Review* 47 (4): 995–1020.

4. Appendices

Annex A. List of essential data items to be collected and processed in Phase II of IMMERSE in compliance with the 'data minimization' principle.

Within the multi-centre, parallel-group cluster randomized controlled trial (cRCT), the intervention will consist of an ESM-based monitoring using the DMMH App in service users and feedback for service users and clinicians on the DMMH dashboard for visualization over a period of 6 months. ESM is a structured diary technique that uses an app to provide service users with short in-the-moment questionnaires on their key problems, symptoms, mood and context as they occur at that specific moment in time. These data thus provide a rich source of information on the current mental state of individuals as well as on patterns of risk and resilience. The ESM-based monitoring using the DMMH App will include a fixed number of items (core items) and ultra-brief additional modules including items that are tailored to the needs of the individual service users. The items are carefully selected to provide the data needed to inform clinical assessment and treatment decisions. Therefore, the data must provide detailed, granular and personalised information about patterns of associations between service users' symptoms, key problem areas and their contexts.

ESM Data (core items)
The structure, content, length and phrasing of the ESM core questionnaire that will be implemented is in line with scientific guidelines (Eisele et al. 2020) and is based on best practices (e.g., Carney et al., 2012; Hanssen et al., 2020; Havermans et al., 2007; Heininga et al., 2019; Littlewood et al., 2018; Maes et al. 2015; Myin-Germeys et al., 2000; Myin-Germeys et al. 2001; Reininghaus et al. 2016).

- Poor sleep quality has been previously found to be related with increased mental health problems the following day, including suicidal ideation (e.g., Littlewood et al., 2018), and an increased intensity of paranoia (Kasanova et al., 2020), identifying sleep as a relevant target variable. While using the DMMH App, participants' **self-reported sleep-quality** will therefore be assessed each morning using 4-5 items (Carney et al., 2012; Littlewood et al., 2018).

- Participants' current **positive affect** (PA) and **negative affect** (NA) will be measured using a selection of 8 items from the Positive Affect Negative Affect Scale (PANAS, Watson et al., 1988) that have been successfully implemented in previous studies (e.g., Myin-Germeys et al., 2000; Myin-Germeys et al., 2001). These data can be used to identify patterns and regularities in participants' mood swings that have been found to be characteristic of poor psychological well-being in general and mood disorders in particular (e.g., Dejonckheere et al., 2019; Houben et al., 2015), and can act as early warning signs for the upcoming onset or offset of depressive episodes (e.g., van de Leemput et al., 2014). Emotional reactivity patterns to daily life-stress have been further found to be related to the experience of psychotic symptoms (e.g., Myin-Germeys et al., 2001), while providing patients with personalized

feedback on their dynamic mood patterns was found to improve symptoms and social functioning in schizophrenia spectrum disorders (Hanssen et al., 2020). Providing insights on the dynamic mood patterns of the individual patient will thus be very relevant to detect and prevent dysfunctional changes in symptoms that are relevant in the treatment of patients with mood-related disorders.

- Information on how symptoms co-vary with contextual factors will be needed to provide information about personalized risk and protective factors that relate to participants' wellbeing and recovery. Therefore, the ESM items will also cover **relevant contextual factors.** These items will focus on how participants experience the present situation (i.e., their momentary **main activity**, as well as potential **social interactions**). The format and content of the items that will be used to assess these contextual information (participants can select their present main activity from a list of categories, which are then followed by further questions, that will depend on their initial response) was successfully used in previous research (Reininghaus et al., 2016). Moreover, this adaptive approach is particularly useful to keep the number of items as short as possible, and thereby reduce the participant burden.

- In addition, **event related stress** experiences will be assessed (using follow-up items referring to the experience of the present situation). The items that have been selected were used in previous research (e.g., Myin-Germeys et al., 2001). An extensive line of research, including the INTERACT study has demonstrated the role of altered stress-reactivity and delayed recovery of stress in daily life, in service users with a variety of mental health disorders (Myin-Germeys et al., 2001; Reininghaus et al., 2016; Vaessen et al., 2019). Which is why stress-reactivity is an important target variable to inform treatment decisions.

- Six items have been selected to cover **emotion regulation strategies** that participants use within a given situation to cope with their emotional and event-related stress experiences. These items have been successfully implemented in previous research (e.g., Barns et al., 2013). Information on the emotion regulation strategies that are used to cope with their present affective experience within a given situation, can reveal to the patient and the clinician whether the strategy used was appropriate and effective, or rather inappropriate and ineffective for the given situation, which can further contribute to a better understanding of these associations and to the definition of highly individualized treatment goals.

- We further aim to integrate a highly individualised category of items that refers to the **treatment goals** of the patient (agreed upon with the therapist). The two most important treatment goals will be entered using short free text fields, followed by two items that capture the participants' **satisfaction with their progress** towards these treatment goals. The structure and content of these items will be informed by an ongoing study that is conducted at the CIMH (DiSERVE@home).

- In addition, three **individual key problem areas** of the participant are defined together with the therapist, which will also be part of the questionnaire used in the DMMH app.

ESM Data (additional modules):
To customise the assessments to the needs of the service user, the therapist may add additional modules to the core item set that are tailored to the symptoms and problems of the service user (e.g., psychotic symptoms, depressive symptoms, experiences related to the medication).

Mobile Sensing Data:
Mobile sensing data will be used by WP 4 to develop new innovative machine learning tools based on multimodal data integration to improve mental health prediction, identify interpretable behavioral contingencies associated with mental health, as well as identify novel mental health predictors. Achieving this goal depends on the data captured by built-in sensors in participants' smartphones being combined with the self-reported ecological momentary measurements (see ESM data). There are multiple advantages from the additional assessment of sensor data: for one, one of the central tenets in the context of multimodal data integration and machine learning is that even features which are weak or non-predictive at the univariate level can become predictive when combined (Koppe et al. 2018), potentially also uncovering novel (higher order) mental health predictors. Moreover, it has been shown that emotional states can be predicted from combinations of sensor data alone with high accuracy (Zhang et al, 2017). Sensor modalities can greatly increase the predictive qualities of mobile apps (Woodward et al. 2019), and they have the advantage that they can be sampled much more frequently and unobtrusively (Seppala et al. 2019), being more robust against false self-reports and ensuring compliance over a longer period of time. They can further help against overfitting models by pooling data (Bzdok et al. 2017). As a minimal requirement for the success of this part of the project, we identified a list of the data that we aim to capture using smartphone sensing features.

Sensor modalities that will be captured have been selected based on evidence that has already shown them to be useful on a univariate level to increase predictability of mental health (and which have already been implemented and assessed via movisens XS).

- **Log App Usage:** Loggs the usage of apps selected. Particularly relevant are applications used for communication (e.g., Facebook Messenger, WhatsApp) and social media (e.g., Instagram, Twitter, Signal). This feature has been successfully used in previous research (e.g. Muehlbauer et al., 2018) to predict changes in the psychopathology of patients with a diagnosis of bipolar disorders (i.e., upcoming manic vs. depressive episodes).
- **Log Phone Calls:** This feature logs the phone activity in an anonymized way (Hashing of phone numbers). By hashing the phone number, first an individual code will be generated. This code together with the phone number will be processed using the SHA1 hash algorithm. Because of that the hash number can not be assigned to the phone number later on. Furthermore it is not possible to compare the hashes of different participants. This feature has been

used to predict participants' symptomatology in previous research (e.g., Ebner-Priemer et al., 2020; Muehlbauer et al., 2018) and will be further relevant to validate the information provided by participants on their social interactions.

- **Log SMS:** This feature logs the SMS activity in an anonymized way (Hashing of phone numbers). Logging incoming and outgoing SMS will provide additional information on participants' activity regarding social interactions, which can provide additional insights on their current mood (e.g., Muehlbauer et al., 2018).
- **Log Location:** This feature logs the location of the participant using GPS, WLAN and Cell. While tracking there are continuous location updates at a maximum rate of every 5s and every 20m. The tracking transitions to a stationary state when the participant remains within 100m of a central position or no location update occurs for 120s. Then more tracking will be done until the participant leaves the 100m radius. This feature has been most regularly used in previous research that has provided strong evidence for its usefulness to identify and predict changes as well as individual differences in participants' mood and symptomatology (e.g., Chow et al., 2017; Friedman et al., 2020; Reichert et al., 2020; Saeb et al., 2015; 2016). Usually, the location parameters are used to examine the activity space (or movement radius) of participants' relative to their home location. The home location can be identified based on the most frequent appearance of the locality of the participant during nighttime (12 am to 6 am) as in previous studies (e.g., Chow et al., 2017; Saeb et al., 2015).
- **Log Physical Activity:** This feature logs the physical activity of the participant (IN_VEHICLE: 0, ON_BICYCLE: 1, ON_FOOT: 2, STILL: 3, UNKNOWN: 4, TILTING: 5). Activity confidence is a value from 0 to 100 indicating the likelihood that the user is performing this activity. Smartphone accelerometers (capturing physical activity) are not very accurate, differ between devices and the wearing position of the Smartphone varies. To accurately measure the physical activity of participants, we might use an additional activity sensor like the movisens Move4. However, logging the physical activity of participants, using the built-in smartphone accelerometers can be useful to draw more robust inferences on their sleep duration, and on their overall physical activity during the day. This feature has been successfully used in previous studies that have not included other external sensors to capture accelerometer data (Muehlbauer et al., 2018; Reichert et al., 2020). For example, participants' activity levels have been found to predict their stress levels, and depressive symptoms (e.g., Ben-Zeev et al., 2015). The data captured using this feature will be most informative when combined with GPS indicators (Log Location).
- **Log Steps:** This feature provides an estimation based on 5 seconds of measurement of the Smartphone accelerometer in a minute. Smartphone accelerometers are not very accurate, differ between devices and the wearing position of the Smartphone varies. To accurately measure steps, an activity sensor like the movisens Move4 would be preferable. However, this feature was used in previous research (e.g., Muehlbauer et al., 2018), and was shown to predict changes in participants' symptomatology (e.g., Ebner-Priemer et al., 2020). The data captured with this feature will be most informative when combined with GPS-based indicators and indicators of physical activity levels.

- **Log Display on/off:** This feature logs whether the display is on or off and provides a measure on how much time participants spend on their phone. This feature has been used in previous research (e.g., Muehlbauer et al., 2018) and can be a useful indicator for participants' behavior when combined with the data captured using other features (e.g., location and app usage).
- **Log Music Listening**: Music plays an important role in emotional regulation (e.g. Miranda et al., 2012), and tracking music listening can be a valuable source of insight as a measure of relaxation/ 'chill-out' time. While tracking whether music is listened to alone could be informative, combining tracking which music is played with machine learning techniques such as embeddings (e.g. Wang et al.,2018) that encodes the emotional context of the music that is listened to could further improve predictiveness over mental health states.

One suggestion to ensure that the data captured using these smartphone sensing features will not be used to identify individual participants is that the incoming raw sensor data is only transferred and stored within the context of WP4 inside the clinical network of the CIMH, and only used for the further algorithm development within WP4, while only completely anonymized higher-level features extracted from the raw data would be used for further processing and visualization in other WP's. This would guarantee that only a small pool of persons has access to the raw data, and that it is solely used in the internal network of the CIMH, where sensitive data is routinely stored and processed (behind the CIMH firewall).

References

Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., & Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric rehabilitation journal*, 38(3), 218–226.

Brans, K., Koval, P., Verduyn, P., Lim, Y. L., & Kuppens, P. (2013). The regulation of negative and positive affect in daily life. *Emotion*, 13(5), 926–939.

Bzdok D, Yeo BT. 2017 Inference in the age of big data: future perspectives on neuroscience. Neuroimage. 2017;155:549–64.

Carney, C. E., Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Krystal, A. D., Lichstein, K. L., & Morin, C. M. (2012). The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep*, 35(2), 287–302.

Chow PI, Fua K, Huang Y, Bonelli W, Xiong H, Barnes LE, Teachman BA. Using Mobile Sensing to Test Clinical Models of Depression, Social Anxiety, State Affect, and Social Isolation Among College Students *J Med Internet Res* 2017;19(3):e62

Dejonckheere E, Mestdagh M, Houben M, Rutten I, Sels L, Kuppens P, Tuerlinckx F. Complex affect dynamics add limited information to the prediction of psychological well-being. *Nat Hum Behav*. 2019 May;3(5):478-491.

Ebner-Priemer, U.W., Mühlbauer, E., Neubauer, A.B. et al. Digital phenotyping: towards replicable findings with comprehensive assessments and integrative models in bipolar disorders. *Int J Bipolar Disord* 8, 35 (2020).

Eisele G, Vachon H, Lafit G, et al. The Effects of Sampling Frequency and Questionnaire Length on Perceived Burden, Compliance, and Careless Responding in Experience Sampling Data in a Student Population. *Assessment*. September 2020.

Friedmann F, Santangelo P, Ebner-Priemer U, Hill H, Neubauer AB, Rausch S, et al. (2020) Life within a limited radius: Investigating activity space in women with a history of child abuse using global positioning system tracking. *PLoS ONE* 15(5): e0232666.

Hanssen, E., Balvert, S., Oorschot, M., Borkelmans, K., van Os, J., Delespaul, P., & Fett, A. K. (2020). An ecological momentary intervention incorporating personalised feedback to improve symptoms and social functioning in schizophrenia spectrum disorders. *Psychiatry Research*, 284, 112695.

Hanssen, E., Balvert, S., Oorschot, M., Borkelmans, K., van Os, J., Delespaul, P., & Fett, A. K. (2020). An ecological momentary intervention incorporating personalised feedback to improve symptoms and social functioning in schizophrenia spectrum disorders. *Psychiatry Research*, 284, 112695.

Havermans, R., Nicolson, N. A., & deVries, M. W. (2007). Daily hassles, uplifts, and time use in individuals with bipolar disorder in remission. *The Journal of Nervous and Mental Disease*, 195(9), 745-751.

Heininga, V. E., Dejonckheere, E., Houben, M., Obbels, J., Sienaert, P., Leroy, B., ... & Kuppens, P. (2019). The dynamical signature of anhedonia in major depressive disorder: positive emotion dynamics, reactivity, and recovery. *BMC Psychiatry*, 19(1), 59.

Houben M, Van Den Noortgate W, Kuppens P. The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychol Bull*. 2015 Jul;141(4):901-30.

Kasanova Z, Hajdúk M, Thewissen V, Myin-Germeys I. Temporal associations between sleep quality and paranoia across the paranoia continuum: An experience sampling study. *Journal of Abnormal Psychology*. 2020 Jan;129(1):122-130.

Koppe et al. 2018., Recurrent Neural Networks in Mobile Sampling and Intervention

Littlewood DL, Kyle SD, Carter L-A, Peters S, Pratt D, Gooding P (2018). Short sleep duration and poor sleep quality predict next-day suicidal ideation: an ecological momentary assessment study. *Psychological Medicine* 49, 403–411.

Maes, I. H., Delespaul, P. A., Peters, M. L., White, M. P., van Horn, Y., Schruers, K., ... & Joore, M. (2015). Measuring health-related quality of life by experiences: the experience sampling method. *Value in Health*, 18(1), 44-51

Markus Reichert, Urs Braun, Sven Lautenbach, Alexander Zipf, Ulrich Ebner-Priemer, Heike Tost, Andreas Meyer-Lindenberg (2020). Studying the impact of built environments on human mental health in everyday life: methodological developments, state-of-the-art and technological frontiers.*Current Opinion in Psychology*, 32, 158-164.

Miranda, D., Gaudreau, P., Debrosse, R., Morizot, J., & Kirmayer, L. J. *Music listening and mental health: Variations on internalizing psychopathology. Music, health, and wellbeing* (pp. 513–529). Oxford University Press.

Mühlbauer, E., Bauer, M., Ebner-Priemer, U. et al. Effectiveness of smartphone-based ambulatory assessment (SBAA-BD) including a predicting system for upcoming episodes in the long-term treatment of patients with bipolar disorders: study protocol for a randomized controlled single-blind trial. *BMC Psychiatry*18, 349 (2018).

Myin-Germeys, I., Delespaul, P. A., & DeVries, M. W. (2000). Schizophrenia patients are more emotionally active than is assumed based on their behavior. *Schizophrenia Bulletin*, 26(4), 847-854.

Myin-Germeys, I., van Os, J., Schwartz, J. E., Stone, A. A., & Delespaul, P. A. (2001). Emotional reactivity to daily life stress in psychosis. *Archives of general psychiatry*, 58(12), 1137-1144.

Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, Mohr DC. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study*J Med Internet Res*2015;17(7):e175

Saeb, S., Lattie, E. G., Schueller, S. M., Kording, K. P., & Mohr, D. C. (2016). The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ*, 4, e2537.

Seppala et al. 2019 Mobile Phone and Wearable Sensor-Based mHealth Approaches for Psychiatric Disorders and Symptoms: Systematic Review

Ulrich Reininghaus, Matthew J. Kempton, Lucia Valmaggia, Tom K. J. Craig, Philippa Garety, Adanna Onyejiaka, Charlotte Gayer-Anderson, Suzanne H. So, Kathryn Hubbard, Stephanie Beards, Paola Dazzan, Carmine Pariante, Valeria Mondelli, Helen L. Fisher, John G. Mills, Wolfgang Viechtbauer, Philip McGuire, Jim van Os, Robin M. Murray, Til Wykes, Inez Myin-Germeys, Craig Morgan, Stress Sensitivity, Aberrant Salience, and Threat Anticipation in Early Psychosis: An Experience Sampling Study, *Schizophrenia Bulletin*, Volume 42, Issue 3, May 2016, Pages 712–722.

Vaessen T, Viechtbauer W, van der Steen Y, Gayer-Anderson C, Kempton MJ, Valmaggia L, McGuire P, Murray R, Garety P, Wykes T, Morgan C, Lataster T, Lataster J, Collip D, Hernaus D, Kasanova Z, Delespaul P, Oorschot M, Claes S, Reininghaus

U, Myin-Germeys I. Recovery from daily-life stressors in early and chronic psychosis. *Schizophr Res*. 2019 Nov;213:32-39.

van de Leemput IA, Wichers M, Cramer AO, Borsboom D, Tuerlinckx F, Kuppens P, van Nes EH, Viechtbauer W, Giltay EJ, Aggen SH, Derom C, Jacobs N, Kendler KS, van der Maas HL, Neale MC, Peeters F, Thiery E, Zachar P, Scheffer M. Critical slowing down as early warning for the onset and termination of depression. *Proc Natl Acad Sci U S A*. 2014 Jan 7;111(1):87-92.

Wang, Dongjing and Deng, Shuiguang and Xu, Guandong (2018). Sequence-Based Context-Aware Music Recommendation.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070.

Woodward et al. 2019. Beyond Mobile Apps: A Survey of Technologies for Mental Well-being.

X. Zhang, W. Li, X. Chen, and S. Lu, "MoodExplorer: Towards Compound Emotion Detection via Smartphone Sensing," Proc. ACM Interact. Mob. *Wearable Ubiquitous Technol*. Article, vol. 1, no. 176, 2017.