IM-TWIN: from Intrinsic Motivations to Transitional Wearable INtelligent companions for autism spectrum disorder

a European funded project

ORDP: Open Research Data Pilot report 1

Deliverable 6.17

Project funded under program H2020-EU.1.2.2.,
Grant agreement ID 952095,
Topic FETPROACT-EIC-06-2019 - EIC Transition to Innovation Activities,
EU contribution € 1 999 965.

Project duration 24 months (November2020, October 2022),
Consortium: Consiglio Nazionale delle Ricerche (ITA), Universiteit Utrecht (NLD), Centre de Recherches Interdisciplinaires (FRA), Università degli Studi di Roma La Sapienza (ITA), Plux-Wireless Biosignals S.A. (PRT).

Deliverable data

Work Package:	6 Management and dissemination
Work Package leader:	CNR
Deliverable beneficiary:	CNR
Dissemination level:	Public
Due date:	31th January (Month 3)
Type:	ORDP: Open Research Data Pilot
Authors:	F. Montedori, V. Sperati, M. Berti[1], G. Baldassarre

Acronyms of partners

CNR-ISTC	Consiglio Nazionale delle Ricerche, Istituto di Scienze e Tecnologie della Cognizione (Italy)
UU	Universiteit Utrecht (The Netherlands)
CRI	Centre de Recherches Interdisciplinaires (France)
LA SAPIENZA	Università degli Studi di Roma La Sapienza (Italy)
PLUX	Plux - Wireless Biosignals S.A. (Portugal)

[1] Author from *Quantum Leap IP*.

Table of contents

1. Overview of the deliverable

In line with the principles of *Open Access* (OA) to scientific publications and research data, generated through H2020 programs, the project IM-TWIN participates in the *Open Research Data Pilot* (ORDP). Accordingly, this deliverable presents the first version of the *Data Management Plan* (DMP), a guidance document for the partners, which establishes the key actions, the methodologies and the strategies for the management of the data that will be generated during the project. The main goal of such policies is to enable the free access, the sharing and the reuse of research data (intended as publications, software, hardware, collected data) by third parties, to promote the exploitation, reproduction and dissemination of the IM-TWIN research outcomes. As good practice, the DMP will then encourage – where possible – the adoption of the FAIR principles[2], to make data *Findable*, *Accessible*, *Interoperable*, and *Reusable*. Importantly, given the nature of the IM-TWIN project, FAIR data must always comply with the requirements – both ethical and due to Intellectual Property Rights IPR – concerning data protection policies.

The DMP structure is inspired to the *Guideline of H2020 programs on FAIR Data Management in Horizon 2020*[3] and includes the following sections:

- **Data Summary**: it reports the type, purpose and the expected format of data;
- **Utility of data**: it describes the target groups – outside the consortium – potentially interested to use the data and the technological outcomes;
- **FAIR data**: it reports the first measures adopted by partners to make the data FAIR (e.g., the online repositories; the publications and licenses policies); limitations to FAIR principles are also reported;
- **Data Security and ethical aspects**: it briefly describes the management of sensitive data which, at least in the early stage of the project, cannot be subject to OA policy (limitations to FAIR principles);
- **Conclusions**: it reports the next deliverables concerning the DMP.

2. Data summary: type, purpose and format

IM-TWIN project is expected to produce the following type of data and technological outcomes:

- scientific publications;
- collections of experimental datasets on human participants;
- Software;
- Hardware.

Additional details on type, purpose and format are provided in the next subsections.

However, as a general rule any data as well as technological outcome of the project, before being disclosed or published, will have to be analyzed to asses the possibility of IPR protection.

[2] https://en.wikipedia.org/wiki/FAIR_data
[3] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

2.1 Scientific publications

The scientific results not subject to IPR protection will be published through an OA policy in international journals and conference proceedings. The published papers will be freely available in the journal's websites, in the dedicated section of the IM-TWIN website and in other free repositories (see sec. 4.1 *Online repositories used by partners* and sec. 4.2 *Publication Policies*).

2.2 Experimental data

During the experimental sessions involving human participants (e.g., children with typical development and diagnosed with Autism Spectrum Disorders – ASD – aged between 12 and 60 months), the following data will be collected and processed:

- **Personal data**: personal information about children and their parents (e.g., age, address, contact details), collected through written informed consent form.
    - format: paper support and corresponding digital support;
    - purpose: mandatory documents in clinical research;
- **Clinical data**: questionnaires and psychological tests for behavioral evaluations.
    - format: paper support and corresponding digital support;
    - purpose: creation of clinical dataset, relevant for the purpose of the research;
- **Biometric data**: video recordings of the experimental activities and physiological data (e.g. Heart Rate, Electrodermal Activity, etc.) collected through wearable sensors.
    - format: standard video files (e.g., mp4, mjpeg, etc.) for video materials; standard files for time series (e.g. CVS, xls, dat, etc.);
    - purpose: creation of dataset to train artificial intelligence systems, based on Machine Learning techniques, to categorize the affective/emotional states of the participants.

2.3 Software

Several software and applications (relying on the technological devices described in the next subsection 2.4 *Hardware*), will be developed and tested during the experimental sessions involving the human participants:

- Software for the analysis of physiological data recorded by wearable sensors.
    - Format: source code (e.g., Python, C++, etc.);
    - Purpose: categorization of the affective states in terms of arousal/valence;
- Software for the analysis of facial features recorded by environmental cameras.
    - Format: (e.g., Python, C++, etc.);
    - Purpose: categorization of the emotional states and facial expressions;
- Software for *Plusme* device.

    - Format: Arduino source code; Apk source code for tablet applications;
    - Purpose: (i) “intelligent” *PlusMe* behaviour, to autonomously stimulate the curiosity and the social interaction in children; (ii) improved interface of the *PlusMe* control tablet, to ease the therapist’s work.

All the aforementioned software has the general purpose to provide important information to therapists and researchers about the effectiveness of the IM-TWIN components for ASD therapy and early ASD diagnosis.

2.4 Hardware

As tangible, technological outcomes, the following devices will be developed or improved for effective use with children:

- Wearable sensors embedded in t-shirts for children aged between 12 and 60 months.
    - format: circuit diagrams and electronic schematics;
    - purpose: non-invasive detection of physiological parameters on children;
- Improved *PulseMe* toy.
    - format: circuit diagrams and electronic schematics;
    - purpose: interactive toy to stimulate curiosity and foster social interaction in children.

3. Utility of data

In the early stage of the project, we identified the main target groups and key actors – described in the following list – potentially interested in adopting and applying both the scientific and the technological outcomes of the project:

- **target group “neurodevelopmental disorders”:** rehabilitation centers; child therapists; child psychiatrists; support teachers; public and private healthcare systems;
- **target group “scientific research”:** researchers in the fields of developmental psychology and computer science, from public and private academic institutions;
- **target group “industrial exploitation”:** small and medium-sized enterprises (SMEs) in the field of hardware, software, medical devices, toys.

4. FAIR data

The next subsections report the first measures adopted by the partners to make the data “FAIR”, namely *Findable*, *Accessible*, *Interoperable*, *Reusable*.

4.1 Online repositories used by partners

In order to ensure the sharing of data, the partners will use the following general-purpose online tools:

- **GitHub**: this is a hosting platform for the development of code, which allows version control, creation and sharing of repositories, collaboration between users with different access level policies. Some repositories for IM-TWIN have been opened by CNR-ISTC at the following link: https://github.com/IM-TWIN.

  It is important to point out that, in the early stage of the project, such repositories will be private: the publication of data will be released only after an IPR analysis;
- **Zenodo**: this is a data archiving tool, developed within the European project Open Aire, used in the H2020 projects participating in the *Open Research Data Pilot*. Zenodo allows the deposit of several types of data as publications, dataset, source codes, and provides useful tools for linking them. Zenodo provides a Digital Object identifier (DOI, a persistent, standardized identifier) for identifying uniquely the archived data, and also features the integration with the abovementioned GitHub[4]. As this is a mandatory requirement, a IM-TWIN page[5] was automatically created by the EU.

Links to GitHub and Zenodo repositories will be also specified in the IM-TWIN website, at the following pages:
www.im-twin.eu/publications
www.im-twin.eu/hardware-and-software

With regard to sensitive data which cannot be subject to OA policies (e.g. video recordings of the experimental sessions with children), partners will exchange data using a secure server at CNR-ISTC (see sec. 5 *Data security*).

4.2 Publications policies

All papers published within the IM-TWIN project, not subject to IPR protection, will be subject to an OA policy, complying with the standard routes:

- **Gold OA**: immediately after the publication, the final version of an article will be freely and permanently accessible for everyone, online on the journal website, with costs sustained by the partners;
- **Green OA**: the final peer-reviewed version of an article (preprint) will be archived in an online repository before, alongside or after its publication (also referred as *self-archiving*). The version that can be deposited into the repository is dependent on the publisher. The repository must be freely accessible for everyone. Repository software usually allows authors to delay access to the article (also referred as *embargo period*).

4.3 Licences policies

Where reasonable, in accordance with IPR evaluations, open data will be published under standard licences[6], ensuring at least authors attributions as for example:

[4] A guide is available here: https://guides.github.com/activities/citable-code/
[5] https://explore.openaire.eu/search/project?projectId=corda__h2020::02f7552bef74ef4e919727a60530c9e0
[6] http://opendefinition.org/licenses/#recommended-conformant-licenses

- Apache Licence 2.0;
- GNU General Public Licence v3.0;
- MIT Licence;
- Creative Commons CC BY 4.0.

4.4 Making data openly accessible

With the exception of scientific publications (which by default follow an OA policy, see sec. 4.2 *Publication policies)*, in the early stage of the project the access policy to data will be subject to several, different restrictions, even between partners. Such limitations are due to the nature of data, essentially sensitive information (as the experimental results about minor participants), and data potentially protectable through Intellectual Property Rights IPR (as innovative software and hardware). A key element for the establishing of access policy is then the management of three requirements, partially in contrast to one other:

- The treatment of sensitive data must always comply with the *General Data Protection Regulation* GDPR 2016/679[7], the European law on privacy and security. This aspect is dealt with more details in the project deliverable 7.3 *Protection of Personal Data POPD, requirement No. 3* (and in general in all the deliverables belonging to the work package 7 *Ethics*);
- The data which can potentially be exploited by the partners, should undergo an IPR evaluation, as described in the project itself;
- All data should be freely available as much as possible, as required by the *Open Research Data Pilot*.

During the project, the above-mentioned requirements will be harmonized, to allow the open access to data as extensively as possible. In more detail the partners, with the advice of *Quantum Leap IP* (a third-party company, expert in IPR strategy) agreed in the early stage of the project to comply with the following policies, according to the type of data:

- **Software:** *Quantum Leap IP* will organize an exchange of views with the partners and industry experts, to evaluate the opportunity to protect proprietary software, be it alone or in combination with hardware. If this is not applicable, the software will be made freely available, according to the rules already described.
- **Hardware:** Same policies applied to software, excluding elements already developed by PLUX: in this case novel solutions, developed throughout the project, will be subject to an in-depth evaluation within the consortium (a detailed IPR analysis is scheduled within the Task 5.4, in the Work Package 5 *Exploitation of IM-TWIN system*).
- **Experimental data:** they will be processed and shared always in accordance with both the requirements by the Ethics Committees and the requirements of European law on privacy, the *General Data Protection Regulation 679/2016.* Detailed information will be

[7] https://gdpr-info.eu/

given in the deliverables compiled in the Work Package 7 *Ethics* (see deliverable 7.3 *Protection of Personal Data POPD, requirement No. 3*).

5. Data security and ethical aspects

As said in the previous sec. 4.1 *Online repositories used by the partners*, the sensitive data (namely the experimental data concerning ASD and TD children) will be exchanged between partners using a secure server at CNR-ISTC. The security measures adopted by CNR-ISTC to manage the access to the server are described in the confidential deliverable 7.3 *Protection of Personal Data POPD, requirement No. 3*, work package 7 *Ethics*.

6. Conclusions

The DMP is a dynamic document: it will be updated alongside the project progress, taking into consideration all the significant changes affecting the data management. Updates on the DMP will be given in the following deliverables:

- D6.18: ORDP, Open Research Data Pilot, report 2, due at M12 (31 Oct 2021);
- D6.19: ORDP, Open Research Data Pilot, report 3, due at M24 (31 Oct 2022).