Managing Affective-learning THrough Intelligent atoms and Smart InteractionS

D11.2 Data Management Action Plan

Workpackage	WP11 - Project Management
Editor(s):	Ana Piñuela, Susana Palomares
Responsible Partner:	ATOS
Quality Reviewers	MaTHiSiS Steering Board (WP Leaders)
Status-Version:	Final – v1.0
Due Date:	30/06/2017
Date:	12/07/2017
EC Distribution:	PU
Abstract:	This Data Management Plan constitutes the updated version of the Data Management Action Plan (DMP) of the MaTHiSiS project and provides a general outline of the project policy for data management.
Keywords:	MaTHiSiS Data Management, Data Collection, Data Documentation, Ethics and legal compliance, Data storage, Data security, Data sharing
Related Deliverable(s)	D11.1 Data Management Action Plan; D2.6 Framework for impact assessment of MaTHiSiS against LEPOSA requirements; D2.7 The impact assessment report, D2.8 Report on monitoring of LEPOSA requirements

This document is issued within the frame and for the purpose of the *MATHISIS* project. This project has received funding from the European Union's Horizon 2020 Programme (H2020-ICT-2015) under Grant Agreement No. 687772

Document History

Version	Date	Change editors	Changes
0.1	07/06/2017	Ana Piñuela, ATOS	Initial TOC and basic contents
0.2	20/06/2017	Ana Piñuela, Susana Palomares, ATOS	Updated contents
0.3	10/07/2017	Ana Piñuela, ATOS	Modifications according to the initial internal revisions
0.4	12/07/2017	Ana Piñuela, ATOS	Modifications according to the final revision
1.0	12/07/2017	Ana Piñuela, ATOS	FINAL VERSION TO BE SUBMITTED

Table of Contents

List of Tables

List of Figures

List of Acronyms

Abbreviation / acronym	Description
ACS	Autism Spectrum Case
AES	Advanced Encryption Standard
API	Application Programming Interface
ATOS	ATOS Spain
CERTH	Centre For Research and Technology Hellas
CGDLC	Career Guidance Distance Learning Case
CLS	Cloud-based Learner's Space
CVS	Concurrent Versions System
DAT	Data Acquisition Tool
DMP	Data Management Plan
FTP	File Transfer Protocol
GDPR	General Data Protection Regulation
IMS	IMS Learning Global Consortium
ITC	Industrial Training Case
IMS LIP	IMS Learner Information Profile
IMS AccLIP	IMS Learner Information Package Accessibility for LIP Information Model
LEPOSA	Legal, Ethical, Privacy, data protection and Social Acceptance
LRS	Learning Record Store
MaTHiSiS	Managing Affective-learning THrough Intelligent atoms and Smart InteractionS
MEC	Mainstream Education Case
NAO	Humanoid Robot
NAS	Network Storage Server
PMLD	Profound and Multiple Learning Disabilities
PMLDC	Profound and Multiple Learning Disabilities Case
SLA	Smart Learning Atoms
T	Task

Abbreviation / acronym	Description
TEL	Technology Enhanced Learning
UM	University of Maastricht
UoN	University of Nottingham
VUB	Vrije Universiteit Brussel
WP3	WP3 Smart Learning Atoms and Learning Graphs
WP4	WP4 Affective and Natural Interaction Instruments
WP5	WP5 Platform Agents
WP6	WP6 Collaboration and Decision Support System
WP7	WP7 System Integration
WP8	WP8 Pilots in Education
WP9	WP9 Industrial Training and Career Guidance Pilots
WP	Work Package
xAPI	ADL Experience API

Table 1: Definitions, Acronyms and Abbreviations

Project Description

MaTHiSiS (Managing Affective-learning THrough Intelligent atoms and Smart InteractionS) is a 36 month duration project co-funded by the European Commission Horizon 2020 Horizon 2020 Programme (H2020-ICT-2015) under Grant Agreement No. 687772. MaTHiSiS consortium is composed of 18 partners from Spain, France, Greece, United Kingdom, Netherland, Belgium, Lithuania and Italy.

MaTHiSiS will create a novel and continuously adaptable "robot/machine/computer"-human interaction educational scheme based on custom-made and adaptable learning goals. The ability of such a system to adapt to different learning requirements and make use of the shared knowledge among its different components will enable new learning methodologies to emerge and foster a new era in learning that goes beyond simple social skill acquisition and targets more workplace-oriented activities.

Non-linearity in learning will be also enhanced through the interaction of the MaTHiSiS ecosystem with learners and tutors in a ubiquitous manner and will incorporate different knowledge and experiences that, once injected to the system, will provide a new, holistic approach by sharing this knowledge across the system components.

One of the core objectives of MaTHiSiS project is to enhance learning environments and make use of computing devices in learning in a more interactive way, which will provide a product-system to be used in formal, non-formal and informal education. An ecosystem for assisting learners/tutors/caregivers for both regular learners and learners with special needs will be introduced and validated in 5 use cases: Autism Spectrum Case (ASC), Profound and Multiple Learning Disabilities Case (PMLDC), Mainstream Education Case (MEC), Industrial Training Case (ITC) and Career Guidance Distance Learning Case (CGDLC).

MaTHiSiS product-system consists of an integrated platform, along with a set of re-usable learning components (educational material, digital educational artefacts, etc.), which will respond to the needs of a future educational framework, and provide capabilities for: i) adaptive learning, ii) automatic feedback, iii) automatic assessment of learner's progress and behavioural state, iv) affective learning and v) game-based learning.

Within MaTHiSiS, an innovative structural tool of learning graphs is going to be introduced to guide the learner through the process of learning in the given scenario. To reach a learning objective, learner will have to "follow the path" of the learning graphs, built up on Smart Learning Atoms. SLAs are atomic and complete pieces of knowledge which can be learned and assessed in a single, short-term iteration, targeting certain problems.

To ensure barrier free integration in the market, MaTHiSiS makes use of a range of interaction devices, such as specialized robots, mobile devices and interactive whiteboards. The consortium ensures easy-to-use solution with e.g. specialized graphical editor-like tool, allowing to easily create educational materials as well as the reusability within both mainstream education and vocational training setups.

A Cloud-based Learner's Space (CLS) will be developed to provide a system for adaptation/personalization in learning, interaction, data acquisition and analysis as well as content creation on the fly. This is a core component of the MaTHiSiS system which includes 3 crucial subsystems which create an innovative smart learning ecosystem: i) the experience engine, a graph-based interactive storytelling engine, that manipulates interactive content that is later sent to a device of tutor's/learner's choice; ii) the learning graph engine, responsible for adaptation of the Learning Graph based on learner's behaviour and interaction; iii) the Decision Support System (DSS) providing and collecting learning analytics and controlling synchronous and asynchronous interaction

between devices. To ensure constant educational flow and augmented learner engagement, the emotion recognition and context aware cognitive/behavioural status extraction tools are introduced within the system addressed by the Sensorial Component.

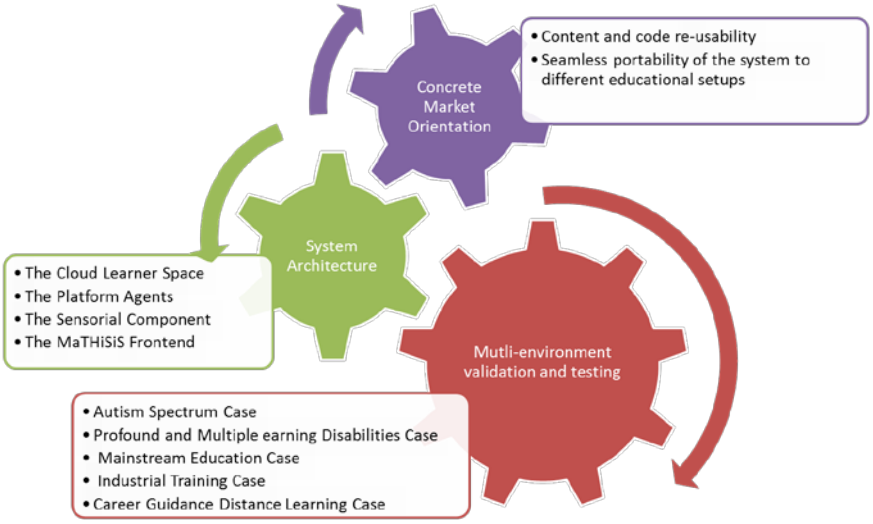

Figure 1: MaTHiSiS in a Nutshell

For the purpose of validating MaTHiSiS approaches in learning environment, a set of Smart Learning Atoms (SLA) is going to be created for defined use cases. Such SLAs will adapt to each learner in a different way based on her/his particular needs, profile, cognitive affective state, relevance to specific learning requirements and previous performance. Further, an editor-like tool is introduced to be able to transform educational material into MaTHiSiS Learning Materials usable by SLAs through Learning Actions. The learning graphs then are going to be deployed to interact with the Cloud-based Learner's Space (CLS) as well as some front-end tools for tutors and caregivers to enable creation, editing and authoring of the learning contents and learning experiences.

MaTHiSiS will support learning across a variety of learning contexts and, with the use of a variety of devices (robots, interactive whiteboards, mobile devices and desktop/laptop computers), with personalized and adaptable, time and location independent learning paths, being transferred between the agents, always taking into consideration best knowledge and practices learnt from the previous device.

By the end of the project, MaTHiSiS will introduce a marketable innovation, aimed at the re-usability of educational and training content and fostering the interactivity between technology and learners/tutors/caregivers.

Executive Summary

This deliverable constitutes the updated version of the Data Management Action Plan (DMP) of the MaTHiSiS project [10] and provides a general outline of the project policy for data management. The overall purpose of this DMP is to support the data management life cycle for all data that will be collected, processed or generated by the project.

The Data management strategy to be follow by MaTHiSiS is be based on the identification and classification of data generated and collected, identification of standards and metadata to be used, exploitation and availability of data as well as how the data will be shared and archived, the preservation of the information; the required ethical, legal compliance and the responsibilities in the implementation of the DMP.

The initial set of the project datasets, namely:

- DATASET 1: Contact users' database
- DATASET 2: Stakeholders' database
- DATASET 3: User requirements' answers
- DATASET 4: Learners' data
- DATASET 5: Learners' interaction data
- DATASET 6: Educational materials
- DATASET 7: Feedback from pilots

This report also includes the description of the sharing, security and ethical mechanisms that guarantee that the personal data elicited and generated during the project is managed and exchanged in compliance with the European Union fair processing principles and the expectations of the persons, the research participants, whose data is shared.

According to the *data sharing protocol*, the personal data of research participants cannot be shared unless the data has been previously pseudonymised. Pseudonymisation, however, is not sufficient, alone, to stem the risks to rights and freedoms that sharing data brings in. Therefore, the following two technical security measures are taken: *1) the encryption of the data and 2) the transmission of the encrypted data over a secured communication network.*

Among the *security measures* to mitigate the risks inherent in data storage and transmission, it can be highlighted the code assigned to each school or institution where the research participant is affiliated, the use of passwords, and the storage of the data acquired in connection with MaTHiSiS in a dedicated computer, in a safe place, lock doors and windows.

Finally, in terms of *ethical and legal compliance aspects*, the participation in the pilots is voluntary and based on the free consent of the person concerned or of his or her legal representative, according to national legislation. Prior to any involvement in the research of MATHISIS all participants have been informed about all the details about MaTHiSiS, the purpose of their participation and how their data will be processed and have signed consent forms.

The monitoring of the implementation of the data sharing, security and ethical/legal protocol are carried out as part of "Task 2.4 Legal, Ethical, Privacy, data protection and Social Acceptance" in "WP2 User, System and Ethics Requirements". All details about the assessment of the activities conducted before the pilots take place are included in the Deliverable 2.8 Report on Monitoring of LEPOSA requirements [1].

The Data Management Plan will be updated over the course of the project whenever significant changes arise.

1. Introduction

This deliverable constitutes the second version of the Data Management Action Plan (DMP) of the MaTHiSiS project and provides an updated version of the project policy for data management, including the following aspects:

- What types of data will the project collect/generate?
- What standards will be used?
- How will this data be shared, stores and transmitted?

This DMP outlines how research data will be handled during the project, and after it is completed. The overall purpose of this DMP is to support the data management life cycle for all data that will be collected, processed or generated by the project. It will contribute to:

- ensure project research data and records are accurate, complete, authentic, interoperable and reliable,
- save time and resources in the long run,
- enhance data security and thereby minimize the risk of data loss,
- ensure research integrity and reproducibility by others,
- prevent duplication of effort by enabling others to use the project's data.

This document is structured as follows:

- Section 2 details the information of MATHISIS datasets listing the eight datasets have been identified.
- Section 3 specifies the specifies the purpose of the data sharing, the circumstances in which the organizations have access to it as well as the data subjects' rights that need to ensured.
- Section 4 outlines the measures foreseen to mitigate the risks inherent in data transmission.
- Section 5 provides a summary of the ethical and legal aspects to be considered in MaTHiSiS.
- Section 6 presents the main conclusions meanwhile the references to the documentation sources used to produce this deliverable are included in the Reference section.

2. MaTHiSiS Data

The general strategy for data management, according to the Guidelines on Data Management in Horizon 2020 [2], will be based on the identification and classification of data generated and collected, standards and metadata to be used, exploitation and availability of data as well as how the data will be shared and archived, the preservation of the information as well as the ethical, legal compliance and the responsibilities in the implementation of the DMP. The MATHISIS DMP will cover all the data life cycle in accordance with the H2020 guidelines regarding Open Research Data [3].

2.1 Data to be collected/ generated

The data to be collected/extracted/generated in MaTHiSiS can be grouped in the following categories:

- Domain knowledge/User related information: Data about Learner profiles, educational materials (learning graphs, SLAs, digital and physical learning materials), quantitative and qualitative data from pedagogical related surveys.
- Sensed raw data: Audio, video, facial expressions gathered from sensorial components, data about the interaction with learning materials.
- Processed data: Information about the affective, cognitive state of learners, recommendations for adaptation and personalization of learning experiences. Structured or unstructured statistics and data about the impact on beneficiaries or targets groups. Structured data about stakeholders for research and dissemination purposes.

2.1.1 Data sources

MATHISIS involves primary data gathering from different sources: 1) semi-structured interviews; 2) stakeholder surveys; 3) Sensors and tracking software mechanisms 4) Communications via MaTHiSiS project website

1. *Semi-structured interviews with stakeholders*: A set of interviews with stakeholders were conducted as part of the activities of "T2.1 User Requirements and learning goals" in "WP2 User, System and Ethics Requirements" with the goal of gathering information about user, technical and legal requirements for the implementation of the MaTHiSiS platform and the tests to be carried out during the pilots ("WP8 Pilots in Education" – "WP9 Pilots in Industrial Training and Career Guidance").
2. *Stakeholder surveys*: Surveys were deployed within WP8 and WP9 to involve participants in the pilots aiming at gathering information about evaluation of the user experience while working with the MaTHiSiS platform during the driver pilots. These surveys will be also distributed during the assisted and real-life pilots.
3. *Sensor and tracking software mechanisms:* They will be used in "WP4 Affective and Natural Interaction Instruments" and "WP6 Collaboration and Decision Support System" to gather information about the learners' (affective and cognitive) status while interacting with the platform during a learning experience, such information will be used to ensure MaTHiSiS capabilities for personalizing and adapting the learning experience that will be developed by WP6.
4. *Communications via MaTHiSiS project website:* The consortium has created a dedicated Website with facilities for bidirectional communications with the users interested in the project: the subscriptions to the project newsletters, the contact form and MaTHiSiS networks of interest. Such functionalities facilitate gathering user feedback that will help to improve MATHISIS platform and services.

2.1.2 Type of data

The type of data to be collected and processed can be either:

- No personal data: such information which is not affected by Data Protection legislation
- Personal data: data which relate to an individual who can be identified
    - (a) from those data, or
    - (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.

In the second case, since MATHISIS follows the European directives, all personal data will be anonymized according to the guidelines derived from activities focussed on ensuring the ethical and privacy issues and compliance to legislation. Such activities are carried out as part of Task 2.4 Legal, Ethical, Privacy, data protection and Social Acceptance in WP2. All details are included in the Deliverable 2.6 [4].

2.2 Standards and metadata to be used

The metadata for the different identified datasets will be generated either automatically by the system or through manual content annotation. It is foreseen that recognised specifications and standards in the Technology Enhanced Learning domain like IMS Learner Information Package specification, xAPI Technical Specification and Learning Resource Metadata Initiative

2.3 Datasets

Currently MaTHiSiS manages seven different datasets that are described below:

2.3.1 DATASET 1: Contact users' database

1	**Dataset reference and name**
	DATASET 1: Contact users' database
2	**Dataset description**
	This set is implemented in Excel and refer to the following sections and purposes: • Contact users' register for newsletter subscriptions, it contains name and e-mail (both mandatory). This dataset is automatically generated when visitors sign up to the newsletter form available on the project website. The register will be used in order to send issues of the project newsletters. • Contact user's personal details with regard to messages sent to the website through the Contact form. It includes name, e-mail, message (all mandatory) and (possibly) phone. The contact details will be used to address the enquiry/request and to send information in the scope of the MATHISIS project.
3	**Standards and metadata**
	This dataset can be imported from, and exported to a CSV, TXT or Excel file.
4	**Data sharing**
	The mailing list will be used for disseminating the project newsletter to a targeted audience. An analysis of newsletter subscribers may be performed in order to assess and improve the overall visibility of the project.

	As this dataset includes personal data, its access is restricted to MATHISIS consortium.
5	**Archiving and preservation (including storage and backup)**
	The dataset will be preserved in the ATOS private infrastructure.

Table 2: DATASET 1: Contact users' database

2.3.2 DATASET 2: Stakeholders' database

1	**Dataset reference and name**
	DATASET 2: Stakeholders' database
2	**Dataset description**
	This dataset is related to the stakeholders that will built the MATHISIS networks (Tutors /Caregivers , Developers, Policy and decision making on mainstream and special education, industrial training and career guidance) and contains the name, surname, job function, domain field/expertise, e-mail contact and location as well as project interests or benefits, what they can contribute, and future actions & communication. This dataset is used to extend and disseminate the information about the project as well as to get feedback from these collaborators. This dataset will not include information about the Learners Networks since all communications with learners will be conducted through their Tutors or Caregivers.
3	**Standards and metadata**
	This dataset can be imported from, and exported to a CSV, TXT or Excel file
4	**Data sharing**
	For these sensitive/restricted data, access restrictions will be enforced (e.g. by requiring specific credentials) or at least limited in the detail available, e.g. by granting an open access exclusively through aggregation, while providing the specific data to authorized users only.
5	**Archiving and preservation (including storage and backup)**
	The dataset will be preserved in the ATOS private infrastructure.

Table 3: DATASET 2: Stakeholders' database

2.3.3 DATASET 3: User requirements' answers

1	**Dataset reference and name**
	DATASET 3: User requirements' answers
2	**Data set description**
	This dataset includes all answers of the stakeholders involved in the pilots (WP8, WP9) about the User requirement gathering carried out as part of WP2 (T2.1) activities.
3	**Standards and metadata**
	Data will be collected and stored using digital audio recording (e.g. MP3) when interviewees consent. The general approach will be to work with transcripts of those interviews accessible .doc file format (Word) and an excel file with the collection of answers per pilots/UC.
4	**Data sharing**

	This dataset will be used to produce reports on the users, system and ethical requirements (WP2) which will serve as reference for the development of MaTHiSiS platform (WP3-WP7). Those reports will only be used by the consortium and its derived User Stories could be shared with external interested parties.
5	**Archiving and preservation (including storage and backup)**
	The dataset will be preserved in the internal project repository (Owncloud)

Table 4: DATASET 3 User requirements' answers

2.3.4 DATASET 4: Learners' data

1	**Dataset reference and name**
	DATASET 4: Learners' data
2	**Dataset description**
	This dataset includes information about the learners' profiles (learning styles, special needs, preferences, learning achievements/performance and current learning goals) and their system user accounts.
3	**Standards and metadata**
	This dataset will be in JSON format. The metadata for this dataset will be described using IMS Learner Information Package specification. Metadata will be defined and processed in XML format according to the different the XML bindings of the specifications/standards used
4	**Data sharing**
	This dataset will be used by all MaTHiSiS components related to the adaptation and personalization (WP4-WP5-WP6). Bearing in mind the private nature of the information included in this dataset it will not be shared with any MaTHiSiS external parties. Only learners will be allowed to recover their data for future purposes.
5	**Archiving and preservation (including storage and backup)**
	The dataset will be preserved in the ATOS private infrastructure.

Table 5: DATASET 4 Learners' data

2.3.5 DATASET 5: Learners' interaction data

1	**Dataset reference and name**
	DATASET 5: Learners' interaction data
2	**Dataset description**
	This data set includes information about the learners' interactions with the MaTHiSiS platform. There will be two categories of data in this set, namely raw sensed data and processed data. The first category includes data gathered by the set of sensors aiming at determining the learners' affective state (facial images, gaze, body skeleton, speech and mobile inertia data); and data about the interaction with learning materials gathered by the platform agents' trackers to determine learner's cognitive state/performance. The second category includes data resulted from processing the first category and will be used for the personalization and adaptation of the learning experience and to show the progress made by the learner towards the achievement of specified learning goal.
3	**Standards and metadata**

	Raw sensed data will include a combination of the following formats (HD video from Kinect 2, including depth and information about head pose, body skeleton, etc.), non-HD video (web camera – also emulating NAO cameras), audio (from Kinect 2, including sound direction) and mobile inertia sensory data (touch, proper acceleration, orientation/rotation, etc.).Meanwhile data about interaction with learning materials will be described according to the xAPI Technical Specification.
4	**Data sharing**
	This dataset will be used by all MaTHiSiS components related to the adaptation and personalization (WP4-WP5-WP6). Bearing in mind the private nature of the information included in this dataset it will not be shared with any MaTHiSiS external parties. Future scientific publications analysing anonymized subsets of dataset will help to disseminate the results of the project in terms of learning process monitoring, adaptation and personalization.
5	**Archiving and preservation (including storage and backup)**
	The dataset will be preserved in the ATOS private infrastructure.

Table 6: DATASET 5 Learners' interaction data

2.3.6 DATASET 6: Educational materials

1	**Dataset reference and name**
	DATASET 6: Educational materials
2	**Dataset description**
	This dataset will include all learning materials to be used in the learning experiences supported by the MaTHiSiS platform as well as the definition of those learning experiences according to the MaTHiSiS concept in the form of learning graphs and SLAs.
3	**Standards and metadata**
	This dataset is a combination of learning materials in different formats (pdf, games, videos, html, etc.) and the learning experiences will be described according to the Learning Action Ontology to be defined in the project. It is expected this set will be also annotated with standardized metadata according to the Learning Resource Metadata Initiative at the end of the project. Metadata will be defined and processed in XML format according to the different the XML bindings of the specifications/standards used
4	**Data sharing**
	This dataset will be mainly used and shared by WP3, WP4, WP5, WP6, which are public. After the last release of the MaTHiSiS platform, all materials and learning experiences will be also shared with external audiences. Also future scientific publications about the MaTHiSiS concept usability and impact will include information about this dataset.
5	**Archiving and preservation (including storage and backup)**
	The dataset will be preserved ATOS private infrastructure. Public access to this dataset will be granted in the Download section of the project website (http://www.MaTHiSiS-

	project.eu/) and through external digital repository (Zenodo, MERLOT or similar).

Table 7: DATASET 6 Educational materials

2.3.7 DATASET 7: Feedback from pilots

1	**Dataset reference and name**
	DATASET 7: Feedback from pilots
2	**Dataset description**
	This dataset will be developed as part of WP8-WP9 activities and will include information about the evaluation of the user experience while working with the MaTHiSiS platform during the pilots, which will serve as input for further refinements of the platform (WP3, WP4, WP5, WP6 and WP7).
3	**Standards and metadata**
	This dataset is a combination of Excel /WORD/PDF documents.
4	**Data sharing**
	This dataset will be shared amongst all project WPs. This dataset will be used to produce reports on the pilots (WP8 and WP9) which will serve as reference for the fine-tuning of the MaTHiSiS platform (WP3-WP7).
5	**Archiving and preservation (including storage and backup)**
	The dataset will be preserved in project internal repository (Owncloud).

Table 8: DATASET 7 Feedback from pilots

3. Data Sharing Protocol

The development of the MaTHiSiS system "requires" the sharing of personal data collected at the pilot sites: the processing is necessary to "train" the technological ecosystem, on the one hand, and to test its applicability in real contexts, on the other. In order to collect the personal data required to attain these two main goals, to train MaTHiSiS system and to test its applicability, five uses cases (Autism, Profound and Multiple Learning Disabilities, Mainstream education, Industrial Training and Career Guidance) or pilots, are organised.

The pilots are organised in different geographical regions, institutions, venues, including schools, companies, and public agencies. Each pilot is divided into four phases: a data acquisition phase, a driver pilot, assisted pilot, and real-life pilot. During each phase, the personal data of participants are collected and exchanged between the partners of MaTHiSiS.

The **data sharing protocol** *specifies the purpose of the data sharing, the circumstances in which the organizations have access to it as well as the data subjects' rights that need to ensured*.

Every organisation involved in MaTHiSiS and in the processing of personal data signs a data sharing protocol. The aim is to ensure that personal data elicited and generated during the project is exchanged in compliance with the European Union fair processing principles and the expectations of the persons, the research participants, whose data is shared.

In MaTHiSiS pilots, as a general rule, the personal data of research participants cannot be shared unless the data has been previously pseudonymised. The **pseudonymisation** of the data of learners is seen as adequate to stem the risk inherent in the sharing of personal data, such as data relating to learning activities, special requirements or needs, and as to avoid that the research participants can be identified. Pseudonymisation, however, is not sufficient, alone, to stem the risks to rights and freedoms that sharing data brings in. Therefore, the following two technical security measures are taken:

- *the encryption of data at the pilot sites*

  Encryption is the process that makes data unintelligible, thereby protecting it against unauthorized access (reading). Good encryption process ensures that the reverse conversion of ciphertexts to the original data is possible only for entities that possess the required cryptographic key (decryption key) [5]. The software used for data compression and encryption in MaTHiSiS is DotNetZip library. DotNetZip is embedded into the Data Acquisition Tool and through the User Interface it allows for compressing and encrypting all data related to the recorded sessions (images, video, database etc). Identical interfaces are integrated to the MaTHiSiS platform in order to provide encryption capabilities for data that will be collected during the Pilots.

- *the transmission of the encrypted data over a secured communication network.*
  The personal data of research participants, after been pseudonymised and encrypted is shared between partners over a secured network. For this purpose, CERTH has setup an FTP Server. An internal firewall protected infrastructure has been configured, containing one physical server acting as FTP Server and a Network Storage Server (NAS). The firewall tunnels requests to specific ports for both upload and download processes. Each partner connects to the FTP server using unique credentials, thus enabling constant monitoring and controlling of user activities. The FTP server is configured to use SFTP, which stands for SSH File Transfer Protocol. In this way the data transferred between remote systems are both encrypted (AES Encyption) and securely transferred. CERTH's infrastructure administration team, monitors constantly the security logs of all servers.

During the data acquisition phase, it may be necessary to adopt a different strategy for the transmission of the data of learners. Such a possibility may arise if the volume of data to be transmitted exceeds what the FTP protocol can support. Companies such as DHL provide transport of external hard drive. This mode of transmission poses specific risks and demands mitigating measures to be adopted. As for transmission via a portable device or external hard drive, these measures must be adopted:

- A record must be maintained detailing the movements and person(s) responsible for, transporting the media;
- The persons responsible must be trained;
- Password protect all portable or remote devices that store personal data;
- Portable that store personal data must be encrypted
- Ensures complete deletion from the portable media, when data arrives at destination.
- Protect the portable media against virus.

The implementation of the data sharing protocol is monitored and documented in periodic reviews-reports in the context of "WP2 User, System and Ethics Requirements". The first assessment is described in the deliverable "D2.8 Report on the Monitoring of the LEPOSA requirements" [1].

4. Data Security Protocol

The *data security protocol* outlines the measures foreseen to mitigate the risks inherent in data storage and transmission.

4.1 Equipment for collection and storage of personal data at pilot sites

During the pilots, several pieces of technology are going to be installed at pilot sites. These include:

1) one PC/laptop. Each partner can decide to set up more than one laptop/PC.
2) USB-keys with data capacity
2) Web cameras as many the PCs/laptops
3) At least one Kinect v2
4) Tripods for the Kinects
5) At least two long (2-5 meters) Ethernet cables.
6) Mobile devices: smartphones or tablets
7) NAO robot (where available).

During the pilots, end user partners store the data captured ("observed" or "generated") by these technologies.

This data is stored locally in the PC/Laptot, or in external hard drive, notably USB-keys with specific data capacities (e.g. 1GB, 5GB of data).

4.2 Obligations and recommendations for END USERS storing personal data of research participants

The following security measures have been agreed with MaTHiSiS partners

To Do's	Recommended / Mandatory	Explanation
Attribute a code to each school or institution where the research participant is affiliated	Mandatory	It should be impossible for anyone apart from the tutor and the end user partner-data controller, to identify the research participant whose data is collected during the pilots.
The personal data of research participants must be pseudonymised.	Mandatory	Pseudonymisation is a technique that consists in replacing one attribute (typically the name) in a record, by another. Only a pseudonymous ID number is used to link individual-level data with participants' identities. The keys to the pseudonymised data are held in in a separate device or room under the responsibility of the pilot's leader. Each partner identifies a researcher responsible for keeping the keys. The researcher must have signed the MaTHiSiS Data Confidentiality agreement (see

To Do's	Recommended / Mandatory	Explanation
		below).
Encryption (I)	Recommended	The pseudonymised data stored in electronic device must be encrypted. This means that personal data of research participants is made unintelligible by transforming it into ciphertext.
Encryption (II)	Mandatory	Thus type of data must be stored in encrypted form : data gathered through sensor technologies regarding; face, gaze, skeleton, speech, inertia ought to be encrypted
Authorised researchers	Mandatory	Only authorized researchers shall have access to personal data. Authorized researchers are researchers regularly employed in partner organizations and who have signed *the MaTHiSiS data confidentiality statement*.
Store data acquired in connection with MaTHiSiS in a dedicated computer, in a safe place, lock doors and windows	Mandatory	This measure reduces the risk of data leaks and unauthorised access
Use a password	Mandatory	Use a password that is strong enough.
Appoint a security officer or a person who will be responsible for the safety of data for the whole duration of the project	Mandatory	This person is the contact person with each pilot site, unless otherwise specified
Keep track of any log in that computer	Recommended	This allows to track authorised access
Delete the data once it has been annotated and transmitted to CERTH.	Mandatory	Please be sure that data has been deleted or irreversibly anonymised

Table 9: Security measures

The implementation of the data security protocol is monitored and documented in periodic reviews-reports in the context of "WP2 User, System and Ethics Requirements". The first assessment is available in "D2.8 Report on the Monitoring of the LEPOSA requirements" [1].

4.3 Specific risks and mitigating measures related to personal data stored on portable devices and transportable media

The table below details a series of risks and mitigating measures that are recommended to protect personal data when it is stored on portable devices and transportable media.

The table has been drafted after the end users partners provided an outline of the security measures in force at their institutions [6]. End user partners are under the obligation to adopt the mandatory risk mitigating strategies.

The source is a security guidance document issued in the USA, in the context of the Health Insurance Portability and Accountability Act (HIPAA) [7]. It must be borne in mind that the guidance applies to storage of medical records, for which the level of security required is the highest possible. However, some of these risks and measures are relevant for MaTHiSiS and hereby adaptively reported.

Risks	Possible risk management strategies	Recommended/Mandatory
Equipment such as Kinect and other devices used in pilot may record personal data or images Data leaks happen when data storage devices are lost or stolen.	Any of the equipment that has the capacity to store personal data must not be left unattended in pilot sites (e.g., in schools), unless data there is are deleted and disposed of.	Mandatory
Portable device is lost or stolen	• Maintain a record of the person(s) responsible for the hardware and electronic media containing personal data; • Password protect all portable or remote devices that store personal data; • Employ encryption technologies of the appropriate strength; • Consider the use of biometrics, such as fingerprint readers, on portable devices.	Compulsory Compulsory Recommended Recommended
Theft of records after inappropriate disposal	Establish personal data deletion policies and media disposal procedures.	Recommended

Table 10: Risks measures

4.4 Retention of data

As a general rule, personal data of research participants is processed when this processing is necessary, that is, required to attain a specific or compatible research purpose.

- *Period of retention for personal data collected at pilot sites during data acquisition phase and the pilot.*

During each phase, researchers and developers working on and in the three layers of the ecosystem are allowed to retain the collected personal data.

At the of the data acquisition and of the pilot, end users shall delete or completely anonymise the personal data of research participants. Picture data must be deleted, unless it is required for further research, e.g., annotations.

- *Period of retention for personal data collected at CERTH, UM, and UoN during data acquisition phase.*

The data collected during the DAT is stored at the premises of CERTH**.**

4.5 Individuals' rights – procedures for dealing with access requests, queries and complaints

Research participants have the right to be informed in a concise, transparent, intelligible and easily accessible form about processing activities concerning him, her or one's child.

Research participants have a right to obtain from the controller confirmation as to whether or not personal data concerning him or her is being processed, and where that is the case, access to the personal data [9].

To exercise their rights as data subjects, research participants must be informed that they can address requests, queries and complaints to the MaTHiSiS partner organization that is closest link to them. These are the user organisations organising the pilots.

Upon receiving the request, the user organization will contact without unreasonable delay and no later than 7 days, the project coordinator (ATOS). The project coordinator will, in cooperation with partner organization VUB, provide a reply to the data subject request, such as a copy of the data collected, to the user organization no later than 15 days from the notification of the request.

4.6 Implementation and review of effectiveness of the sharing agreement

The data storage system at the premises of the technical coordinator (CERTH) will monitor the access, the privileges of the users, the uploads and downloads requests to the system.

During the pilot phases, it may be necessary to store data in the cloud. In this case, the task of monitoring performed by CERTH will be taken over by ATOS.

CERTH is also responsible for the secure transmission of personal data a collected in the pilots.

4.7 Sanctions for failure to comply with the agreement or breaches by individual staff

A failure to comply with the requirements of this protocol entails the suspension of the authorization to access personal data on the researcher and, if applicable, depending on the seriousness of the breach, on the organization partner.

Criminal sanctions remain applicable under different jurisdictions for disclosing personal information outside the project.

5. Ethics and legal compliance aspects

All details about ethics and legal compliance in terms of Current EU legislative initiatives, anonymisation procedures, consent needed, restrictions on 3rd parties, procedures for handling sensitive data and data owners are included in WP2 deliverables related to ethical issues, in particular "D2.6 Framework for impact assessment of MaTHiSiS against LEPOSA requirements" and "D2.7 The impact assessment report"[8]. A short summary of the MaTHiSis protocol followed in MaTHiSiS is included below.

Participation in the pilots is voluntary and based on the free consent of the person concerned or of his or her legal representative, according to national legislation.

As for the **information requirements**, prior to any involvement in the research of MATHISIS, participants have been informed with regards to the following:

1. An overall description of the MATHISIS system;
2. The purpose of the experiment/pilot;
3. The experiment/ pilot procedures;
4. The responsibility of the participant;
5. Reasonable foreseeable risks and inconveniences as well as benefits;
6. Voluntary nature of the participation in the experiment/pilot (consent can be withdrawn at any time without consequence);
7. Foreseeable circumstances and reasons where the participant's involvement may be terminated;
8. Type and extent of data collected; categories and exact types of data collected; purpose of collection for each data type; amounts of data collected;
9. Confidentiality of information collected; how/where/for how long it will be stored; security measures in place; who will have access;
10. Confidentiality of participant identity (publication of results will be as part of a group average);
11. Expected duration of the experiment/pilot and of the participant's specific participation;
12. Approximate number of participants in the experiment/pilot;
13. Incidental findings policy.

As far as personal data protection requirements are concerned, each participant or their legal representative, if applicable, will be informed about:

1. The type(s) of data to be collected;
2. The method(s) of collecting data;
3. Confidentiality and anonymity conditions associated with the data including any exceptions to confidentiality, for example, with respect to potential disclosures and details on possible sharing of personal information with authorized third parties on a strict need-to-know basis;
4. The opportunity to have any supplied data destroyed on request (up to a specified date);
5. The name and contact details of the person(s) responsible for the data collection and processing;
6. How the data will be processed and disclosed. See below on perosnal data protection.

All participants will sign **consent forms**. A sample of informed consent form is presented below:

Informed consent form for participation in research

I, undersigned [name] [date and place of birth – natural person; registry number – legal entities] [contact details] [if representing a minor: her/his name, date of birth, etc.], hereby give my consent to take part in the research carried out by the MaTHiSiS Research Consortium.

1. I have been informed that the MaTHiSiS project (*Managing Affective-learning THrough Intelligent atoms and Smart InteractionS*) is a research project currently run under the Horizon 2020 Framework Programme under the grant agreement no. [number]. The co-ordinator of the project is [name] [registry number – legal entities] [contact details], who might be contacted with regard to any question regarding my participation.
2. I have been informed about the purposes of the project. I have had all my questions answered to my satisfaction.
3. My participation in the research will include [describe in detail].
4. Information obtained during the research will be used for [describe in detail].
5. My personal data will be made available only to the members of the MATHISIS Consortium.
6. I understand that any further use of this information will require my separate consent.
7. [I require not to be identifiable in any research results.]
8. I understand [I will not] be paid for my participation.
9. I give this consent fully informed, freely and voluntarily and I understand that I am free to withdraw my consent and discontinue my participation at anytime without any negative consequences.
10. The relevant laws of [country] shall apply.

Done in two copies, of which one is for the MaTHiSiS Consortium and one for the participant.

Done at [place] on [date].

Signature

Protection of the personal data of research participants

MaTHiSiS partner VUB monitors the activities and verifies effective compliance with European and Partners undertaking research (directly or indirectly) in connection with personal data stemming from the pilots, are aware of their obligations under the Data Protection Directive (95/46/EC) and the General Data Protection Regulation (GDPR) as well as national data protection laws, national privacy and data protection laws.

Guiding principles

The processing of personal data of research participants under the MaTHiSiS project are based on the following general principles:

- *Necessity*: all personal data collection and processing operations are organised so as to minimise the use of personal and identification data.
- *Purpose*: personal data and processing operations are acquired and performed exclusively for the purposes of MaTHiSiS project.

- *Fairness*: the data is collected, stored and transmitted exclusively to test MaTHiSiS technology.
- *Safety*: Personal data is kept safe and is accessible only by authorized personnel;
- *Transparency*: those concerned by the processing of their personal data have the right to know the rational behind them and have data subjects rights

The implementation of the recommendations is monitored, and documented in periodic reviews-reports. This ensures that the LEPOSA impact assessment remains transparent. At this respect, "D2.8 Report on monitoring of LEPOSA requirements covers the activities conducted in between December 2016 and June 2017", the period between the submission of Deliverable 2.7 "The LEPOSA report" and its first review. The assessment is publicly accessible thus the public at large can be informed about the existence of an assessment process, its terms of reference, the method and its progress.

6. Conclusion

This document is an updated version of D11.1 Data Management Action Plan. It provides an updated overview of the data that MATHISIS project will manage during its lifetime together with the definition of procedures implemented by MATHISIS project to efficiently manage its research data in terms of sharing, security and ethics.

The report on the monitoring of the activities related to the implementation of the data sharing, security and ethical protocols conducted before the driver pilots took place in May 2017 is available in “D2.8 Report on monitoring of LEPOSA requirements”. The assessment is publicly accessible thus the public at large can be informed about the existence of an assessment process, its terms of reference, the method and its progress.

7. References

[1] VUB (eds). D2.8 Report on monitoring of LEPOSA requirements. Deliverable of the MaTHiSiS project, 2017

[2] Guidelines on Data Management in Horizon 2020, V2.0, 30 October 2015, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf .

[3] Guidelines on Open Access to Scientific Publication and Research Data in Horizon 2020, Version 2.0, 30, October 2015, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

[4] VUB (eds). D2.6. Framework for impact assessment of MaTHiSiS against LEPOSA requirements (The LEPOSA report). Deliverable of the MATHISIS project, 2016

[5] ENISA, Recommended cryptographic measures *Securing personal data 2013* p.17

[6] See Ethical clearance MS3. List of Annexes.

[7] HIPAA Security Guidance, 2006. https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/securityrule/remoteuse.pdf

[8] VUB (eds). D2.7. The impact assessment report. Deliverable of the MATHISIS project, 2016.

[9] Art. 15 GDPR, As the CJEU clarified in the Rijkeboer case, the right to access is necessary to enable the data subject to exercise his rights under art. 12 (b).

[10] ATOS (eds). D11.1 Data Management Action Plan. Deliverable of the MaTHiSiS project, 2016.