

Managing Affective-learning THrough Intelligent atoms and Smart Interactions

D11.1 Data Management Action Plan

Workpackage	WP11 - Project Management
Editor(s):	Carmen L. Padrón-Nápoles
Responsible Partner:	ATOS
Quality Reviewers	Lazaros Gymnopoulos (CERTH) Andrew Pomazanskyi (NG)
Status-Version:	Final
Date:	15/12/2016
EC Distribution:	RE
Abstract:	This Data Management Plan includes the identification and classification of data generated and collected in MaTHiSiS, the set 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 ethical, legal compliance and the responsibilities in the implementation of the DMP.
Keywords:	MaTHiSiS Data Management, Data Collection, Data Documentation, Ethics and legal compliance, Data storage and Backup, Data Selection and preservation, Data sharing
Related Deliverable(s)	D11.2 Data Management Action Plan; D2.6 Framework for impact assessment of MaTHiSiS against LEPOSA requirements; D2.7 The impact assessment report



Document History

Version	Date	Change editors	Changes
0.1	20/05/2016	Carmen L. Padrón Nápoles, ATOS	Initial TOC and basic contents
0.2	07/07/2016	Carmen L. Padrón Nápoles, ATOS	Modifications according to the initial internal revisions
0.3	14/12/2016	Carmen L. Padrón Nápoles, ATOS	Modifications according to the internal revisions
1.0	15/12/2016	Carmen L. Padrón Nápoles, ATOS	FINAL VERSION TO BE SUBMITTED

The information and views set out in this document are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Table of Contents

1. Introduction	9
1.1 Document structure	9
2. Data Management strategy	10
2.1 Data to be collected/ generated	10
2.1.1 Data sources	10
2.1.2 Type of data	11
2.2 Standards and metadata to be used	11
2.3 Exploitation, availability of data and re-use	11
2.4 Archiving and preservation	11
2.5 Ethics and legal compliance	11
2.6 Responsibility and resources	12
3. Project Datasets	13
3.1 DATASET 1: Contact users' database	13
3.2 DATASET 2: Stakeholders' database	14
3.3 DATASET 3: User requirements' interviews	14
3.4 DATASET 4: User requirements' answers	15
3.5 DATASET 5: Learners' data	15
3.6 DATASET 6: Learners' interaction data	16
3.7 DATASET 7: Learning materials	17
3.8 DATASET 8: Feedback from pilots	17
4. Conclusion	19
5. References	20

List of Tables

<i>Table 1: Definitions, Acronyms and Abbreviations.....</i>	<i>5</i>
<i>Table 2: DATASET 1: Contact users' database.....</i>	<i>14</i>
<i>Table 3: DATASET 2: Stakeholders' database.....</i>	<i>14</i>
<i>Table 4: DATASET 3 User requirements' interviews</i>	<i>15</i>
<i>Table 5: DATASET 4 User requirements' answers</i>	<i>15</i>
<i>Table 6: DATASET 5 Learners' data</i>	<i>16</i>
<i>Table 7: DATASET 6 Learners' interaction data.....</i>	<i>17</i>
<i>Table 8: DATASET 7 Learning materials</i>	<i>17</i>
<i>Table 9: DATASET 8 Feedback from pilots.....</i>	<i>18</i>

List of Acronyms

Abbreviation / acronym	Description
ACS	Autism Spectrum Case
API	Application Programming Interface
ATOS	ATOS Spain
AV	Aerospace Valley
CERTH	Centre For Research and Technology Hellas
CGDLC	Career Guidance Distance Learning Case
CLS	Cloud-based Learner's Space
D	Deliverable
DMP	Data Management Plan
DoA	Description of the Action
DXT	DIGINEXT
EOPPEP	National Organisation for the Certification of Qualifications and Vocational Guidance
FMD	Fondazione Mondo Digitale
IMOTEC	Institute of Mobile Technologies for Education and Culture
IMS	IMS Learning Global Consortium
IPR	Intellectual Property Rights
ITC	Industrial Training Case
IMS LIP	IMS Learner Information Profile
IMS AcCLIP	IMS Learner Information Package Accessibility for LIP Information Model
IWB	Interactive whiteboard
JCYL	Consejería de Educación_Junta de Castilla y León
LCS	La Cometa del Sud
LEPOSA	Legal, Ethical, Privacy, data protection and Social Acceptance
LRS	Learning Record Store
LG	Learning Graphs
LGlib	Learning Graph implementation API

Abbreviation / acronym	Description
LPR	Learner Profile Repository
M	Month
MaTHiSiS	Managing Affective-learning THrough Intelligent atoms and Smart InteractionS
MEC	Mainstream Education Case
NAO	Humanoid Robot
NCSR	National Center For Scientific Research "Demokritos"
NG	Nurogames
NTU	Nottingham Trent University
O.S	Operating System
OTE	Hellenic Telecommunications Organisation S.A.
PA	Platform Agent
PE	Polo Europa
PMLD	Profound and Multiple Learning Disabilities
PMLDC	Profound and Multiple Learning Disabilities Case
RIX	RIX Research and Media
SC	Sensorial Component
SLA	Smart Learning Atoms
T	Task
TEL	Technology Enhanced Learning
UC	Use Case
UM	University of Maastricht
UoN	University of Nottingham
VUB	Vrije Universiteit Brussel
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 problems¹.

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

¹ Smart Learning Atom (SLA) definition in the DoA MaTHiSiS – Part B, page 8

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.

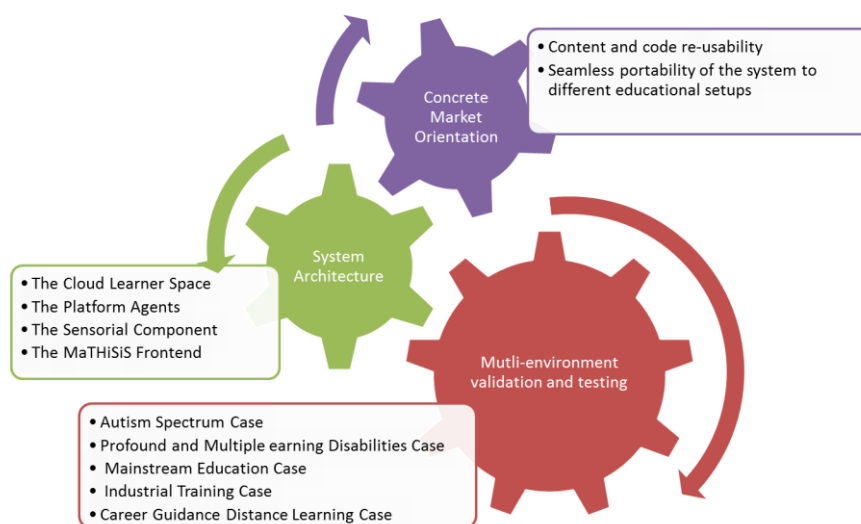


Figure 1 MaTHiSiS in a Nutshell

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 initial version of the Data Management Action Plan (DMP) of the MaTHiSiS project 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 followed by MaTHiSiS is 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' interviews
- DATASET 4: User requirements' answers
- DATASET 5: Learners' data
- DATASET 6: Learners' interaction data
- DATASET 7: Learning materials
- DATASET 8: Feedback from pilots

Has been identified and described according to the aspects of such strategy.

This document reflects the current state of consortium agreements regarding data management and is consistent with those referring to exploitation and protection of results. It is a living document that is expected to mature during the project lifetime and will be updated accordingly.

The next version of the MATHiSiS Research Data Management Plan will put a strong emphasis on the complete definition of procedures to be implemented by MATHiSiS project to efficiently manage its research data in terms of storage and backup (backup provision, recovery procedure), selection and preservation (which data will be retained/shared/ preserved, length of time data to be preserved and preservation preparation time). It will also include the description of the mechanisms for systematic anonymization of personal data that will be working before the first stage of pilots will start in M14.

1. Introduction

This deliverable constitutes the initial version of the Data Management Action Plan (DMP) of the MaTHiSiS project and provides a general outline 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 exploited and/or shared/made accessible for verification and re-use? Reasons why the data cannot be made available in some cases.

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.

The described policy herein reflects the current state of consortium agreements regarding data management and is consistent with those referring to exploitation and protection of results. It is a living document that is expected to mature during the project lifetime and will be updated accordingly.

1.1 Document structure

This document provides an initial description of the Data Management Action Plan (DMP) within the MATHiSiS project through the following sections:

- Section 2 defines the Data management strategy to be followed during the MATHiSiS project according to the Guidelines on Data Management in Horizon 2020 [1] published by the EC services.
- Section 3 details the information of MATHiSiS datasets: listing the eight datasets have been identified, at this early stage of the project., the main exploitation perspectives for each of those datasets, and the major management principles the project will implement to handle those datasets.
- Section 4 presents the main conclusions meanwhile the references to the documentation sources used to produce this deliverable are included in the Reference section.

2. Data Management strategy

The general strategy for data management, according to the Guidelines on Data Management in Horizon 2020 [1], 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 [2].

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) focus group discussions and stakeholder workshops; 3) Sensors and tracking software mechanisms 4) Communications via MaTHiSiS project website and, 5) External content repositories

1. *Semi-structured interviews with stakeholders*: A set of interviews with stakeholders will be conducted as part of activities of T2.1 in WP2 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-WP9).
2. *Focus group discussions and stakeholder workshops*: Focus groups and workshops conducted within WP8 and WP9 will involve participants in the pilots aiming at gathering information about evaluation of the user experience while working with the MaTHiSiS platform during the testing periods. Furthermore such workshops will be used to collect information about the economic impact of MaTHiSiS and a set of transferable lessons for further refinements of the platform and the learning experiences supported by it.
3. *Sensor and tracking software mechanisms*: They will be used in WP4 and WP6 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.

5. *External content repositories:* Different open learning repositories will be used to retrieve learning materials that will be used when designing learning experiences according to MaTHiSiS concept (learning graphs, SLAs) and during the deployment of such learning experiences with the MaTHiSiS platform.

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 [3].

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 Exploitation, availability of data and re-use

MATHISIS project intends to make its non-personal/sensitive research data available for open access and use by other researchers. In furtherance of this goal, MATHISIS will deposit its research data into a digital repository ([Zenodo](#) or similar).

For sensitive/restricted data access restrictions will be enforced (e.g. by requiring specific credentials, anonymization) 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.

2.4 Archiving and preservation

All research data will be stored and backed up regularly through existing back-up mechanisms in place at ATOS internal infrastructures. Qualitative data will be backed up and secured by the coordinator on a regular basis and metadata will include clear labelling of versions and dates..

2.5 Ethics and legal compliance

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 will be included in the deliverables D2.6 and D2.7. Meanwhile information about the expected embargo periods and, detailed data about MaTHiSiS licenses ([Creative Commons Attribution Licenses](#) and others) will be included in future updates of this document as part of the D11.2 Data Management Action plan to be delivered in M18.

Additionally, the Grant Agreement and the MATHiSiS Consortium Agreement are to be referred to for further details on the ownership, management of intellectual property and access.

2.6 Responsibility and resources

ATOS, as project coordinator and leader of T11.4 Data Management Plan, CERTH as project technical coordinator and VUB as leader of Legal, Ethical, Privacy, data protection are responsible for the DMP implementation.

All storage resources will be located at ATOS internal infrastructure and all partners will be involved in the different DMP activities according to their role in project as described in the MaTHiSiS DoA.

3. Project Datasets

The current version of the DMP addresses the following aspects on a dataset by dataset basis and presents the current status of reflection within the consortium in relation to the set of data that will be managed by the project.

- Dataset reference, name and reference: Identifier for the data set to be gathered/processed/generated. Given the date of delivering this initial version of the DMP no references to the datasets will be included. This information will be updated in the next version of the DMP to be submitted in M18.
- Dataset description: Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.
- Standards and metadata: it includes reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.
- Data sharing: A detailed description of how data will be shared, including access procedures, outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).

In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, and commercial, privacy-related, security-related).

- Archiving and preservation: Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

Currently MaTHiSiS plans to manage 8 different datasets. Next sections describe those datasets following an structure in accordance with the Guide of Horizon 2020 for the Data Management Plan [1].

3.1 DATASET 1: Contact users' database

1 Dataset reference and name

DATASET 1: Contact users' database

Dataset Reference: *Link will be provided at a later stage of the MATHISIS project*

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

3.2 DATASET 2: Stakeholders' database

1	Dataset reference and name
	DATASET 2: Stakeholders' database Dataset Reference: <i>Link will be provided at a later stage of the MATHISIS project</i>
2	Dataset description
	This dataset is related to the stakeholders that will build the MATHISIS networks (Tutors /Caregivers and Developers) 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

3.3 DATASET 3: User requirements' interviews

1	Dataset reference and name
	DATASET 3: User requirements' interviews Dataset Reference: <i>Link will be provided at a later stage of the MATHISIS project</i>
2	Dataset description
	This dataset will contain the set of questionnaires to be prepared by WP2 in order to gather information from stakeholders involved in the pilots (WP8, WP9) to define the initial set of users and system requirements, which will serve as reference for the development of MaTHiSiS platform and the definition of the pilots' plans.
3	Standards and metadata
	This dataset can be imported from, and exported to a CSV, TXT or Excel file
4	Data sharing
	This dataset will be used to produce reports on the users, system and ethical requirements which will serve as reference for the development of MaTHiSiS platform (WP3-WP7) and the planning of pilots (WP8-WP9). Those reports will only be used by the consortium
5	Archiving and preservation (including storage and backup)m.
	The dataset will be preserved in the internal project repository (Owncloud)

Table 4: DATASET 3 User requirements' interviews

3.4 DATASET 4: User requirements' answers

1	Dataset reference and name
	DATASET 4: User requirements' answers Dataset Reference: <i>Link will be provided at a later stage of the MATHISIS project</i>
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 5: DATASET 4 User requirements' answers

3.5 DATASET 5: Learners' data

1	Dataset reference and name
	DATASET 5: Learners' data Dataset Reference: <i>Link will be provided at a later stage of the MATHiSiS project</i>
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 won't 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 6: DATASET 5 Learners' data

3.6 DATASET 6: Learners' interaction data

1	Dataset reference and name
	DATASET 6: Learners' interaction data Dataset Reference: <i>Link will be provided at a later stage of the MATHiSiS project</i>
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; 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 won't 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 7: DATASET 6 Learners' interaction data

3.7 DATASET 7: Learning materials

1	Dataset reference and name
	DATASET 7: Educational materials Dataset Reference: <i>Link will be provided at a later stage of the MATHISIS project</i>
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 8: DATASET 7 Learning materials

3.8 DATASET 8: Feedback from pilots

1	Dataset reference and name
	DATASET 7: Feedback from pilots Dataset Reference: <i>Link will be provided at a later stage of the MATHISIS project</i>
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). The set will also include reports about the economic impact of MaTHiSiS and the set of transferable lessons

	according to the results of the pilots. (WP1)
3	Standards and metadata
	This dataset is a combination of Excel /WORD/PDF documents.
4	Data sharing
	This dataset will be mainly shared amongst all project WPs. Only the reports about economic impact and the set of transferable lessons will be shared with external audiences as part of the project's future scientific publications .
5	Archiving and preservation (including storage and backup)
	The dataset will be preserved in project internal repository (Owncloud), Public access to sharable subset of this dataset will be granted in the Download section of the project website (http://www.MaTHiSiS-project.eu/) and through external digital repository (Zenodo or other)

Table 9: DATASET 8 Feedback from pilots

4. Conclusion

This deliverable provides an initial overview of the data that MATHISIS project will manage during its lifetime together with challenges and constraints that need to be considered to ensure project's research data and records will be accurate, complete, interoperable and reliable; to enhance data security and minimize the data loss risks; to prevent duplication of efforts allowing others to use some of the data managed by the project.

The next version of the MATHISIS Research Data Management Plan will put a strong emphasis on the complete definition of procedures to be implemented by MATHISIS project to efficiently manage its research data in terms of storage and backup (backup provision, recovery procedure), selection and preservation (which data will be retained/shared/ preserved, length of time data to be preserved and preservation preparation time). Special attention will be given to ensuring that all generated datasets do not infringe either partner IPR rules or regulations related to personal data protection. A clear and complete mechanism for systematic anonymization of personal data will be defined and is planned to be in place before the first stage of pilots will start.

5. References

- [1] 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.
- [2] 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.
- [3] Eugenio Mantovani and István Böröcz (eds). D2.6. Framework for impact assessment of MaTHiSiS against LEPOSA requirements (The LEPOSA report). Deliverable of the MATHISIS project, 2016