D6.2 – Data Management Plan

Project No	GA824160
Project Acronym	EnTimeMent
Project full title	ENtrainment & synchronization at multiple TIME scales in the MENTal foundations of expressive gesture
Instrument	FET Proactive
Type of action	RIA
Start Date of project	1 January 2019
Duration	48 months

Distribution level	[PU][1]
Due date of deliverable	Month 6
Actual submission date	June 2019
Deliverable number	6.2
Deliverable title	Data Management Plan
Type	R
Status & version	
Number of pages	
WP contributing to the deliverable	6
WP / Task responsible	UNIGE
Other contributors	All partners
Author(s)	VBC
EC Project Officer	Teresa De Martino
Keywords	Data management

[1] **PU** = Public, **PP** = Restricted to other programme participants (including the Commission Services), **RE** = Restricted to a group specified by the consortium (including the Commission Services), **CO** = Confidential, only for members of the consortium (including the Commission Services).

Contents

Abbreviations

DMP	Data Management Plan
CA	Consortium Agreement
EC	European Commission
EU	European Union
GA	Grant Agreement
OA	Open access
WP	Work Package

1 Summary

This report focuses on the preparation of the Data Management Plan (DMP) for EnTimeMent project. DMP provides an analysis of the main elements of data management policy that will be used throughout the project with regard to all datasets that will be generated. In particular, DMP will define how this data will be managed and shared by the project partners, and also, how this information will be curated during the project, as well as preserved after the project ends.

DMP of EnTimeMent project describes the life cycle of all modelling and observation data collected and processed during the project, giving an overview of available research data, access and data management as well as terms of use. The DMP reflects the current state of the discussions, plans and ambitions of the partners and will be updated and augmented with new datasets and results during the lifespan of EnTimeMEnt project. The first version of the DMP is delivered in Month 6 of the project, when the first data sets are identified. More detailed version of the DMP will be delivered at later stages of the project if applicable and at the end of the project as a final DMP report.

2 Introduction

The amount of data generated is continuously increasing, while use and re-use of data to derive new scientific findings is relatively stable. This information would be useful in the future if the data is well documented according to accepted and trusted standards which enable the recognition of suitable data by negotiated agreements on standards, quality level and sharing practices. For this purpose, DMP defines strategies to preserve and store data over the defined period of time in order to ensure their availability and re-usability after the end of EnTimeMent project. The procedures that will be implemented for data collection, storage, access, sharing policies, protection, retention and destruction will be according to the requirements of the national legislation of each partner and in line with the EU standards.

2.1 Building a DMP in the context of H2020

The EC provided a document with guidelines for project. The guidelines address aspects such as research data quality, sharing and security. Following these guidelines, DMP will be developed with aim to provide a consolidated plan for EnTimeMent partners in the data management plan policy that the project will follow. The consortium will comply with the requirements of Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regards to the processing of personal data and on the free sharing of such data. The consortium will preserve the right to privacy and confidentiality of data of the survey participants, by providing them two documents: The Participant Information Sheet and the Consent Form. These documents will be sent electronically and will provide information about how the answers will be used and what the purpose of the survey is. The participants will be assured that their answers will be used only for the

purposes of the specific survey. The voluntary character of participation will be stated explicitly in the Consent Form. Before conducting the survey, the consortium will examine and follow the requirements of the national legislation in line with the EU standards, whether the proposed data collection requires special local/national ethical/legal permission.

An ethical approach will be adopted and maintained throughout the fieldwork process. The responsible partners will assure that the EU standards regarding ethics and Data Management are fulfilled. Each partner will proceed with the survey according to the provisions of the national legislation that are adjusted in line with the respective EU Directives for Data Management and ethics. The consortium will follow a transparent recruitment process for the engagement of stakeholders and inclusion/exclusion criteria for all the surveys will be explained in the Participant Information Sheet.

The Steering Committee of the project will also ensure that EU standards are followed. The issue of informed consent for all survey procedures, all participants will be provided with a Participant Information Sheet and Consent Form to provide informed consent. The default position for all data relating to residents and staff will be anonymous.

3 EnTimeMent Data Management Plan

3.1 Description of Work

The purpose of this document is to set the DMP for the EnTimeMent project. It contains guidelines that will be used for the development of a DMP which will include an analysis of the main elements of the data management policy that will be used by the EnTimeMent consortium with regards to all the data that will be generated by the project. Moreover, the DMP will cover the following aspects:

- Description of the data to be collected / created
- Standards / methodologies for data collection and management
- Ethics and Intellectual Property concerns or restrictions
- Plans for data sharing and access
- Strategy for long-term preservation

The DMP will not be a fixed document, but it will evolve and will gain more precision and substance during project implementation. New versions of the DMP will be created whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors.

3.2 Purpose of the Data Management Plan

The DMP purpose is to:

• Support the data management life cycle for all data that will be collected, processed or generated by the Project.

• Provide an analysis of the main elements of the data management policy which will be used by the applicants with regard to all the datasets which will be generated by the Project.

• Provide detail and guarantee about the preservation of the data collected during the Project, as well as any results derived from the associated research.

• Provide detail on how the EnTimeMent consortium plans to address the Ethical issues (if any) related to data which will be collected during the Project timeframe.

• Create a document which explains the management of data collected during the Project.

The DMP is not a fixed document, but it will evolve during the Project.

Also, the DMP will be oriented to:

- EnTimeMent project's participant organisations
- Local Ethics Committee
- Partners' personnel and all stakeholders interested in the Project
- European Commission

Finally, the DMP contains details on

- Brief description of data types which will be collected during the EnTimeMent project, explaining the procedures used to collect or create them.
- Copyright and IPR issues.
- Ethical issues related to data storage, persons authorised to see/use them and how long they are kept; managing ethical concerns that include the anonymisation of data; procedures used to obtain the consent requested to allow data sharing and reuse.

4 Data Sets

4.1 Define Research Data

One definition of research data is: "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings." [2]. Research data covers a broad range of types of information, and digital data can be structured and stored in a variety of file formats. Note that properly managing data (and records) does not necessarily equate to sharing or publishing that data.

Some examples of research data include:

- Documents (text, Word), spreadsheets
- Laboratory notebooks, field notebooks, diaries
- Questionnaires, transcripts, codebooks
- Audiotapes, videotapes
- Photographs, films
- Spectra
- Test responses
- Slides, artefacts, specimens, samples
- Collection of digital objects acquired and generated during the process of research
- Database contents (video, audio, text, images)
- Models, algorithms, scripts
- Contents of an application (input, output, logfiles for analysis software, simulation software, schemas)
- Methodologies and workflows
- Standard operating procedures and protocols

In addition to the other records to manage, some kinds of data may not be sharable due to the nature of the records themselves, or to ethical and privacy concerns. As defined by the OMB [1], this refers to:

- Preliminary analyses
- Drafts of scientific papers
- Plans for future research
- Peer reviews
- Communications with colleagues

Research data also do not include:

[2] Office of Management and Budget (OMB), Uniform Administrative Requirements for Grants and Agreements With Institutions of Higher Education, Hospitals, and Other Non-Profit Organisations, CIRCULAR A-110 REVISED 11/19/93 As Further Amended 9/30/99**,** https://www.whitehouse.gov/omb/circulars_a110#36

- Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published or similar information which is protected under law.
- Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.

The following research records may also be important to manage during and beyond the life of a project:

- Correspondence (electronic mail and paper-based correspondence)
- Project files
- Grant applications
- Ethics applications
- Technical reports
- Research reports
- Signed consent forms

4.2 Data Sets

The specific Data Sets for the EnTimeMent project need to be identified and described with the contribution of all project partners. A short description of the data, which will be generated in the research project (e.g. samples, physical collections, software, curriculum materials, and other materials to be produced during the course of the project) must be provided. Additionally, an estimation of the amount of data and content of the data (if possible) must be included.

For this reason, the following template will be filled by the tasks leaders in order to collect information regarding data sets according to the following template.

TASK (Mention the Task or Tasks this dataset belongs)	
Data Identification	
Dataset description	(Description, Source of data, creation of data)
Source	(Data derives from…)
Partners activities and responsibilities	
Partner owner dataset	

Partner in charge of data collection	
Partner in charge of data analysis	
Partner in charge of data storage	
Standards	
Info about metadata	
Standards, Format	(standard if any, word, excel, design etc.)
Data exploitation and sharing	
Data exploitation	(purpose/use of the data analysis)
Data sharing, re-use and distribution	(Data shared with…, Use of data by…)
Archiving and preservation (including storage and backup)	
Data storage (including backup)	Storage and backups of the relevant materials … first level of storage and backup. e.g. Google Drive folder - second level of storage A third level of storage and accessibility will be the members section in the project website (Private documents).

In addition to the above, the following table will also be filled by the tasks leaders in order to provide adequate information regarding data sets according to the following issues:

Are you generating the data or sourcing it from somewhere else under certain terms and conditions?
Is the data digital or non-digital, or both?

How will the data be created or collected? What instruments or tools will be used to produce the data?
What transformations will the data undergo? What software or file formats will you use as you work with the data?
Will the data be updated or become redundant as you make revisions and produce subsequent versions?
Is the data sensitive or confidential?
Is there ethics approval, or is ethics approval required?

From the information that will be gathered the roles of the partners and the use of the data will be identified. As a result, for each type of research data it will be defined who will be providing the data and who will be using/analysing the data.

Additionally, the file formats that will be used are an important issue. The formats that will be used should be the best for long-term preservation and continued access of data. Formats most likely to be accessible in the future are:

- Non-proprietary and not tied to a specific piece of software
- Open, documented standard
- Common, used by the research community
- Standard representation (ASCII, Unicode)
- Unencrypted
- Uncompressed

4.3 Descriptive information and Metadata

The DMP defines what documentation and metadata will accompany the data. Metadata is structured information describing the characteristics of a resource; for example, the dates associated with a dataset or the title and author of a book. Metadata supports discovery, re-use and long-term preservation of resources. Metadata needs to vary across scientific fields, but typically cover the following:

- General descriptive and access of metadata
- Data characteristics
- Archive terms and access policies

A metadata record consists of a set of predefined elements that define specific attributes of a resource. Each element can have one or more values; for example, a dataset may have multiple creators.

Documenting data enables other researchers to discover your data. Metadata about the nature of the files is also critical to the proper management of digital resources over time.

All the partners will agree on specific issues regarding for example:

- The way that the data will be organized or formatted so that everyone working on it now and in the future knows the origins of the data.
- The way that each file will be named (File Naming Conventions). The use of the following format is proposed for each file/document: "Date (yyyymmdd)_project_company_filename_author_version".
- Providing adequate metadata within the dataset (e.g. field labels or column headings) in order to be easy to interpret the data. Other examples of information that the data need to contain include:
  - Reference period
  - Project funding information: European Union logo and information about Grant Agreement and the action/program that funds the project
  - Release policy including dissemination rules and purposes
  - Information about data collection (source, frequency and adjustments)
  - Keywords (Keywords or phrases describing the subject or content of the data)
  - Geographic coverage of the dataset (if applicable)
  - File formats
  - Comments
- Ways to identify different versions. It is proposed in each data set to include a versioning table, additionally to use the prefix ".v1" in each file/document name relevant to the versioning table. For versioning the rule that will be followed will be the use of a sequentially numbered system: v1, v2, v3, etc and "Final" for the final version. If changes need to be done in the final version then the name of the document will change including the relevant sequential version number, ensuring that the document with the "Final" prefix is indeed the final one.

At a minimum, metadata records should be kept in a fielded form, such as a spreadsheet, CSV file, or tab-delimited file. Auxiliary information necessary to interpret the metadata - such as explanations of codes, abbreviations, or algorithms used - should be included as accompanying documentation.

A brief description of the Data sets identified for the EnTimeMent project are included in Annex 1.

4.4 Ownership (IPR)

In the DMP issues regarding copyright and Intellectual Property Rights of the data are included. These issues are set in the Consortium Agreement and the Grant Agreement of the EnTimeMent project regarding all the results of the project. Thus, the DMP follows the Consortium Agreement and the Grant Agreement that is signed by all project partners regarding Ownership issues.

Any type of results generated under the EnTimeMent Project will be disseminated in accordance with the Consortium Agreement.

4.5 Storage and Access

The long-term storage of data is a main issue of the project. Not only the data will be stored. The computers, the recording tools should be stored in order to confirm, validate, reproduce the experimentations.

To prevent accidental destruction, the storing centers should be at least two, placed in two different geographical places that are at least 30 km far from each other. The data access as well as all other procedures should be identical for both storing centers.

One must distinguish two classes of data types that need different storages:

a. Storage of data in their initial format. Typically, one finds here video data, big size data (more than 1 Go), compressed data whose size is as big and the decompressing is irregular;

b. Data whose storage can be created with a software (e.g. *.csv files in excel). It is wise to save these data in a database. Remind that the storage with these tools should save the way to reconstruct the initial data. A typical situation is when one import to a new version of excel an old file and this import can alter the initial data. Unfortunately, in these cases one should save the old version of the tool and the operating system that works with it. Another strategy is to convert data to an open format like CSV, JSON, XML, etc. This permits to transfer data without any dependence to a software. In order to reduce the risk of data destruction or corruption and the delays of data transfer one has to reduce the number of formats of data. More precisely, the CSV format is relevant for table-oriented data with a limited number of columns (less than 1000 columns) and with a much more important number of rows. In practice, it happens to cut a CSV file into little files that are easier to import for other treatments. Of course, these is relevant for full matrices. The sparse tables need to use sparse format techniques that are well developed for numerical purposes (see below).
c. The sparse case needs a more pertinent data format, JSON. This format organizes data in a tree like structure. This format permits to model relation structures like “is a part of”, “is a” and so on. For example: { 'year':{'quarter1':{ 'jannuary':31, 'february':[28,29], 'mars':31} } }. Beware that this format reduces the need for memory by increasing the complexity to access information. This access needs a tree search and is not completely standard for the moment. A notation of type DOM could be a solution to this issue as many tools allow the utilization and visualization of this data type and the queries in this format.

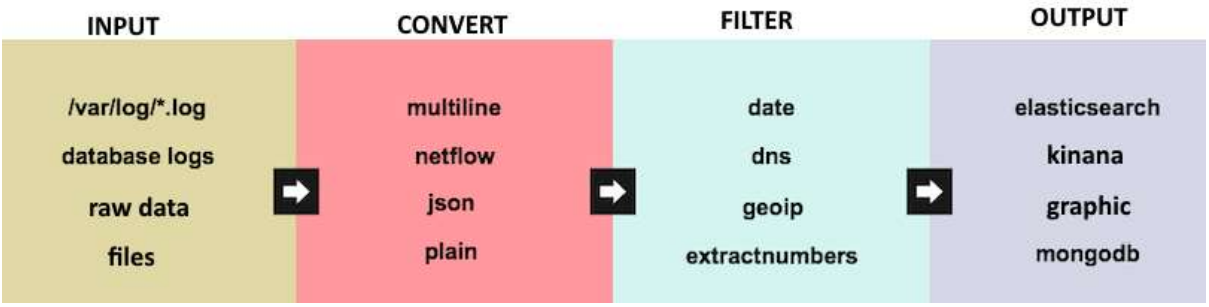
Classically, the more similar structures with these two types of data formats are the data bases. It is a routine to import a CSV file to a SQL database of type MySql, MariaDB or Postgresql. In fact, a similarity in format makes it easy, if not automatic, to import / export these data formats. The complications come sometimes from the « very rich » structure of data. This structure makes it difficult to include these data into a column of a database. It can be extremely difficult to insert an image or a sound in a database format (except in a binary format but than one loses the capacity to query information in a database way). Moreover, database need periodic updates and recording that permit to retrieve data when a technical problem happens. This is a situation analogous to excel versions and old files. On have to ensure the export of data to an open standard (e.g. SQL-92) and verify that the export of data is possible through a test off-line that simulates a crash of the database.

Remains the issue of importing into these databases non-tabular data like the JSON format. In these cases, the use of new types of databases NoSQL (No=NotOnly SQL) permits to override the format and then record a hierarchical structure of the information. A technical description of this type of database will be given later as this formalism seems interesting to EnTimeMent project. One must take care on the merging phase of the trees of JSON files in order to avoid data destruction or overlap.

A last format, not mentioned until now, is the link format of graph type. There is a wide variety of formats that describe binary graphs as well as hypergraphs. The recording formats are not yet stabilized (sparql, graphql, …). These formats manage problems like insufficient return of data (under-fetching)

or over-fetching. The graphical format of data is not widely used as it complexify the projects. Whenever needed, this format could be carefully considered as it can be the cornerstone of a project and all information should be organized as a function of the graphical format. About this last point, it appears that a tabular / relational structure of information goes in SQL database in order to record temporal data (time-series, eeg acquisitions, …). Next, a NoSQL database is added for data with undefined format or for data with a strong hierarchical structure that don't allow to create a chain of related tables. It is estimated that for a number of links greater than 40/50 the structure of information is very complex to be queried in a coherent way by database tools. Actually, MySql limits the queries to a limit of 60 datalinks.

Adding a non-relational database will give a richer information, e.g. extraction to a software stack ElasticSearch / LogStash / Kinana.



About the treatment's reproducibility issue, EuroMov develops its expert knowledge about a tool that chains the treatments by grouping sets of data. This tool (actually under development) is based on the galaxy project (https://galaxyproject.github.io/). Its programming in a Docker image will permit a simplified deployment of treatments.

This solution permits to replay the data treatment in a uniformized manner. On the other side it needs to learn how to use it to a high level in order to enrich it and interact with galaxy. The solutions of type stack of softwares Elasticsearch /Logstash / Kinana are under development by the community of software engineers.

About specific formats of movement data, up to know data are temporal observations. The c3d format (https://www.c3d.org) is the standard in motion capture data. A set of existing software tools some of which are opensource, permits to interact and manipulate this type of data.

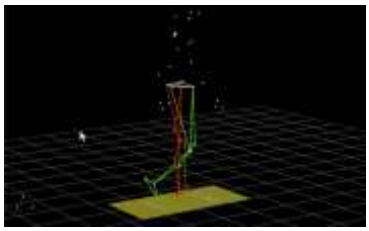


Figure 1biomechanical sample viewer

When recording these types of data, their format depends often on the capture hardware. The next step is the data acquisition soft. For example, the data acquisition toolbox of Matlab creates a format daq file. It is then possible to convert this format with Matlab or otherwise open and treat these files with other programming languages. We suggest saving daq files but for further use the data should be transformed into other formats like CSV or TSV (Tab Separated Values).

The Coordinator of the EnTimeMent project along with the Dissemination & Exploitation Manager will be in charge for data management and all the relevant issues.

5 Ethical and Legal issues

5.1 Ethical Issues

The EnTimeMent partners are to comply with the ethical principles as set out in Article 34 of the Grant Agreement, which, among other, states that all activities must be carried out in compliance with:

(a) Ethical principles (including the highest standards of research integrity — as set out, for instance, in the European Code of Conduct for Research Integrity — and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct)

(b) Applicable international, EU and national law

Specifically, the EnTimeMent project consortium handle Ethics and legal issues following a strictly defined set of procedures that have been described in a specific deliverable (**D7.1**). We point to that document for a complete description.

The operational procedures employed for the design of science and technology will be guided by ethical concerns. The project will foster adherence to recognised ethical practices as highlighted in the General Principles and Requirements applicable to Researchers in the European Charter for Researchers. Fundamental care will be given to the determination of how to best include ethics considerations as a guidance for technological and research developments, considering potential usages and misuses.

5.2 Confidentiality

All EnTimeMent partners must keep any data, documents or other material confidential during the implementation of the project and for four years after end of the project in accordance with Article 36 of the Grant Agreement. Further detail on confidentiality can be found in Article 36 of the Grant Agreement. In addition, all partners will comply with the General Data Protection Regulation 2016/679. Under the regulation, the data controllers and processors are fully accountable for the data processing operations. Any violation of the data subject rights may lead to sanctions as described in Chapter VIII, art.77-84.

6 Conclusions

The document presented the Data Management and Open Access strategy for the EnTimeMent project. The initial EnTimeMent Data Management Plan comprises the identification of the initial Data Sets by all partners. The DMP will be revised and updated during the entire duration of the project. The DMP will be updated at least by the mid-term and final review to fine-tune it to the data generated and the uses identified by the consortium since not all data or potential uses are clear from the start. New

versions of the DMP will be created, whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors.

7 Annexes

7.1 Annex 1. EnTimeMent Data Sets

7.1.1 WP1

WP1 Theoretical Foundations, Computational Models, and Algorithms	WP1 data sets will be further defined and amended later in to the project

7.1.2 WP2

WP2 Experiments	
Tasks **2.1, 2.2, 2.3**	
Data Identification	
Dataset description	The following data will be collected: - information that identifies the individual and their personal characteristics (such as name, surname date and place of birth, fiscal code); - motion capture data, video recordings and electrophysiological data (electromyography, electroencephalography).
Source	Human participants
Partners activities and responsibilities	
Partner owner dataset	IIT
Partner in charge of data collection	IIT
Partner in charge of data analysis	IIT
Partner in charge of data storage	IIT
Standards	

Info about metadata	TBD
Standards, Format	C3D for motion capture, .AVI for video recordings, matlab matrices for electrophysiological data
Data exploitation and sharing	
Data exploitation	Investigate the neurobehavioral basis of movement organization and perception at multiple time scales during behavioural interaction and coordination
Data sharing, re-use and distribution	Where participants give permission, motion capture, video recordings and neurophysiological data will be shared with EnTimeMENT partners and the wider research community
Archiving and preservation (including storage and backup)	
Data storage (including backup)	All data will be stored on encrypted hard disks in locked cabinets in IIT.

In addition to the above table, the following table will also provide adequate information regarding the data sets for WP2.

Are you generating the data or sourcing it from somewhere else under certain terms and conditions? Data will be newly captured.
Is the data digital or non-digital, or both? Digital
How will the data be created or collected? What instruments or tools will be used to produce the data? Motion capture sensors will be used to capture body movement data. Videos will be captured via a full HD camcorder. Neurophysiological data will captured by appropriate sensors (e.g., surface electromyography sensors).
What transformations will the data undergo? What software or file formats will you use as you work with the data? We will analyse the data using various custom written analytical tools.
Will the data be updated or become redundant as you make revisions and produce subsequent versions? Data collected will not be revised, but will be added to as more participants are recorded.

Is the data sensitive or confidential? Sensitive or confidential data such as names, surnames, fiscal code, date and place of birth will be store on an internal database accessible only by the experimenter and password locked.
Is there ethics approval, or is ethics approval required? Ethics approval for the collection of the data has been received from local Research Ethics Committee and will be extended to include the specific requirements of the project (as detailed in WP7.1).

For the WP2 there are additional data sets from the UNIGE as shown in the table below

WP2 Experiments	
Tasks **2.2, 2.3**	
Data Identification	
Dataset description	The dataset will consist of body movement data, video recordings, audio recordings, and electrophysiological (?) signals.
Source	Human participants
Partners activities and responsibilities	
Partner owner dataset	UNIGE
Partner in charge of data collection	UNIGE
Partner in charge of data analysis	UNIGE
Partner in charge of data storage	UNIGE
Standards	
Info about metadata	TBD
Standards, Format	C3D for motion capture, mpeg4 for videos, AIFF for audios, matlab matrices for electrophysiological signals
Data exploitation and sharing	
Data exploitation	Investigate the basis of movement analysis and prediction at multiple time scales during behavioural interaction and coordination

Data sharing, re-use and distribution	Where participants give permission, motion capture and multimodal data can be shared with EnTimeMENT partners and the wider research community
Archiving and preservation (including storage and backup)	
Data storage (including backup)	All data will be stored on encrypted hard disks in locked cabinets at UNIGE.

Are you generating the data or sourcing it from somewhere else under certain terms and conditions? Data will be newly captured.
Is the data digital or non-digital, or both? Digital
How will the data be created or collected? What instruments or tools will be used to produce the data? Motion capture and multimodal sensors will be used to capture body movement data.
What transformations will the data undergo? What software or file formats will you use as you work with the data? We will analyse the data using various custom written analytical tools.
Will the data be updated or become redundant as you make revisions and produce subsequent versions? Data collected will not be revised, but will be added to as more participants are recorded.
Is the data sensitive or confidential? No data will be sensitive or confidential.
Is there ethics approval, or is ethics approval required? Ethics approval for the collection of the data has been received from local Research Ethics Committee and will be extended to include the specific requirements of the project (as detailed in WP7.1).

In addition, for the WP2 there are further data sets from the UM as shown in the table below

WP2 Experiments
Tasks **2.1, 2.2, 2.3**
Data Identification

Dataset description	A dataset consisting of human fMRI measurements and stimuli of human body movements (video)
Source	Human participants
Partners activities and responsibilities	
Partner owner dataset	UM
Partner in charge of data collection	UM
Partner in charge of data analysis	UM
Partner in charge of data storage	UM
Standards	
Info about metadata	tbd
Standards, Format	DICOM and NIFTII for fMRI data, mpeg4 for videos
Data exploitation and sharing	
Data exploitation	The relation between brain activity and body movement parameters at multiple time scales
Data sharing, re-use and distribution	Participants sign a informed consent on data sharing. When signed positive, the base fMRI data and brain maps can be shared with the consortium and the wider community.
Archiving and preservation (including storage and backup)	
Data storage (including backup)	Data will be stored on UM servers specifically design to store and protect this kind of data.

Are you generating the data or sourcing it from somewhere else under certain terms and conditions? Data is generated
Is the data digital or non-digital, or both? Digital

How will the data be created or collected? What instruments or tools will be used to produce the data? This will be fMRI data, generated by an MRI.
What transformations will the data undergo? What software or file formats will you use as you work with the data? We will analyse the data using various custom written analytical tools.
Will the data be updated or become redundant as you make revisions and produce subsequent versions? Data will not become redundant, once recoded the data cannot be updated.
Is the data sensitive or confidential? No data will be sensitive or confidential.
Is there ethics approval, or is ethics approval required? Ethics approval for the collection of the data has been received from local Research Ethics Committee and will be extended to include the specific requirements of the project (as detailed in WP7.1).

In addition, for the WP2 there are further data sets from the EuroMov as shown in the table below

WP2 Experiments	
Tasks **2.1, 2.2, 2.3**	
Data Identification	
Dataset description	Human posture and movements, along with physiological data
Source	Human participants
Partners activities and responsibilities	
Partner owner dataset	EuroMov
Partner in charge of data collection	EuroMov
Partner in charge of data analysis	EuroMov
Partner in charge of data storage	EuroMov
Standards	
Info about metadata	tbd
Standards, Format	CSV or C3D

Data exploitation and sharing	
Data exploitation	We address the relation between movements in solo, duo or group and at multiple time scales
Data sharing, re-use and distribution	Any dataset can be shared on demand, for explicit research purpose. Once anonymized, part of the data might be widely distributed, for example on OpenData servers (case by case analysis is necessary).
Archiving and preservation (including storage and backup)	
Data storage (including backup)	Storage is on EuroMov dedicated and secured computers, with backup on the EuroMov server.

Are you generating the data or sourcing it from somewhere else under certain terms and conditions? EuroMov generates data
Is the data digital or non-digital, or both? Digital
How will the data be created or collected? What instruments or tools will be used to produce the data? We measure movement with VICON mocap, eventually associated to other physiological data (ECG, EMG, EEG, gazing, breathing, and specific data acquisition tools as needed).
What transformations will the data undergo? What software or file formats will you use as you work with the data? We analyse the data using custom written software.
Will the data be updated or become redundant as you make revisions and produce subsequent versions? Original raw data cannot be modified. The data processing pipeline (data curation, pre-processing) includes partly redundant datasets.
Is the data sensitive or confidential? Personal data is always sensitive and confidential. Once anonymized, pre-processed data might no longer be confidential (case by case analysis is necessary).
Is there ethics approval, or is ethics approval required? EuroMov always obtain ethics approval before recording human data for research purposes.

7.1.3 WP3 & WP4

WP3 Data Acquisition and Multi-Time Signal Analysis and Processing **WP4** Use Case Scenarios	
Tasks **3.1, 3.2, 3.7, 3.9 & 4.2**	
Data Identification	
Dataset description	The dataset will consist of body movement data and mental experiences captured in individual movement settings and dyad movement settings, with and without musical sonification.
Source	Human participants
Partners activities and responsibilities	
Partner owner dataset	UCL
Partner in charge of data collection	UCL
Partner in charge of data analysis	UCL
Partner in charge of data storage	UCL
Standards	
Info about metadata	TBD
Standards, Format	Video, motion capture data (2D/3D joint angles/positions in digital text format), self-reported data (handwritten text, typed text, audio, or other digital form), physiological data where possible (various forms)
Data exploitation and sharing	
Data exploitation	To address the following research questions: 1. How can movement data be auto-segmented at multiple timescales? What temporal segments of movement (from multiple timescales) map to different mental experiences, and how can these segments be auto-mapped to these labels? Can we create musical trajectory that reflects action trajectory, such that motives occur at movement segment boundaries?

	2. To what extent does a person's movement behaviour change, based on a partner's performance of the same movement type, and at what timescales do the changes occur? To what extent can musical sonification generated be synchronised to encourage improved movement quality for the person, through sonically-supported entrainment?
Data sharing, re-use and distribution	Where participants give permission, motion capture, physiological data, and self-reported data can be shared with EnTimeMENT partners and the wider research community.
Archiving and preservation (including storage and backup)	
Data storage (including backup)	Video data will be stored on encrypted hard disks in locked cabinets in UCL. Motion capture, physiological data, and self-reported data will be stored on UCL OneDrive.

In addition to the above table, the following table will also provide adequate information regarding the data sets for WP3 & WP4.

Are you generating the data or sourcing it from somewhere else under certain terms and conditions? Data will be newly captured.
Is the data digital or non-digital, or both? Both.
How will the data be created or collected? What instruments or tools will be used to produce the data? Motion capture sensors and videos, where possible, will be used to capture body movement data, and mental experiences will be captured using self-report. When physiological data is captured, appropriate sensors (e.g. surface electromyography sensors) will also be used.
What transformations will the data undergo? What software or file formats will you use as you work with the data? We will analyse the data using various machine learning and statistical analysis software. Audio data will be transcribed, and text data will be digitised.
Will the data be updated or become redundant as you make revisions and produce subsequent versions? Data collected will not be revised, but will be added to as more participants are recorded.
Is the data sensitive or confidential?

Video data will be confidential. Further, un-anonymised self-report will be sensitive.
Is there ethics approval, or is ethics approval required? Ethics approval for the collection of the data has been received from the UCL Research Ethics Committee.

Also, for WP3 there are additional data set from UNIGE as depicted in the table below along with the adequate information regarding the specific data sets

WP3 Data Acquisition and Multi-Time Signal Analysis and Processing	
Tasks **3.1, 3.3, 3.4, 3.6, 3.7, 3.8, 3.9**	
Data Identification	
Dataset description	The dataset will consist of body movement data, video recordings, audio recordings, and electrophysiological (?) signals.
Source	Human participants
Partners activities and responsibilities	
Partner owner dataset	UNIGE
Partner in charge of data collection	UNIGE
Partner in charge of data analysis	UNIGE
Partner in charge of data storage	UNIGE
Standards	
Info about metadata	TBD
Standards, Format	C3D for motion capture, mpeg4 for videos, AIFF for audios, matlab matrices for electrophysiological signals
Data exploitation and sharing	
Data exploitation	Investigate the basis of movement analysis and prediction at multiple time scales during behavioural interaction and coordination
Data sharing, re-use and distribution	Where participants give permission, motion capture and multimodal data can be shared with

	EnTimeMENT partners and the wider research community
Archiving and preservation (including storage and backup)	
Data storage (including backup)	All data will be stored on encrypted hard disks in locked cabinets at UNIGE.

Are you generating the data or sourcing it from somewhere else under certain terms and conditions? Data will be newly captured.
Is the data digital or non-digital, or both? Digital
How will the data be created or collected? What instruments or tools will be used to produce the data? Motion capture and multimodal sensors will be used to capture body movement data.
What transformations will the data undergo? What software or file formats will you use as you work with the data? We will analyse the data using various custom written analytical tools.
Will the data be updated or become redundant as you make revisions and produce subsequent versions? Data collected will not be revised, but will be added to as more participants are recorded.
Is the data sensitive or confidential? No data will be sensitive or confidential.
Is there ethics approval, or is ethics approval required? Ethics approval for the collection of the data has been received from local Research Ethics Committee and will be extended to include the specific requirements of the project (as detailed in WP7.1).

In addition, for WP3 & WP4 there are further data set from UM as depicted in the table below along with the adequate information.

WP3 Data Acquisition and Multi-Time Signal Analysis and Processing **WP4** Use Case Scenarios
Tasks **3.1, 3.2, 4.4**
Data Identification

Dataset description	A dataset consisting of human fMRI measurements and stimuli of human body movements (video)
Source	Human participants
Partners activities and responsibilities	
Partner owner dataset	UM
Partner in charge of data collection	UM
Partner in charge of data analysis	UM
Partner in charge of data storage	UM
Standards	
Info about metadata	TBD
Standards, Format	DICOM and NIFTII for fMRI data, mpeg4 for videos
Data exploitation and sharing	
Data exploitation	The relation between brain activity and body movement parameters at multiple time scales
Data sharing, re-use and distribution	Participants sign an informed consent on data sharing. When signed positive, the base fMRI data and brain maps can be shared with the consortium and the wider community.
Archiving and preservation (including storage and backup)	
Data storage (including backup)	Data will be stored on UM servers specifically design to store and protect this kind of data.

Are you generating the data or sourcing it from somewhere else under certain terms and conditions? Data is generated
Is the data digital or non-digital, or both? Digital

How will the data be created or collected? What instruments or tools will be used to produce the data? This will be fMRI data, generated by an MRI.
What transformations will the data undergo? What software or file formats will you use as you work with the data? We will analyse the data using various custom written analytical tools.
Will the data be updated or become redundant as you make revisions and produce subsequent versions? Data will not become redundant, once recoded the data can not be updated.
Is the data sensitive or confidential? No data will be sensitive or confidential.
Is there ethics approval, or is ethics approval required? Ethics approval for the collection of the data has been received from local Research Ethics Committee and will be extended to include the specific requirements of the project (as detailed in WP7.1).

Moreover, for WP3 there are further data set from DU as depicted in the table below along with the adequate information.

WP3 Data Acquisition and Multi-Time Signal Analysis and Processing	
Tasks **3.1, 3.7**	
Data Identification	
Dataset description	A dataset consisting of audiovisual, motion capture and biosignal recordings of adults carrying out individual and group music-related tasks
Source	Human participants
Partners activities and responsibilities	
Partner owner dataset	DU
Partner in charge of data collection	DU
Partner in charge of data analysis	DU
Partner in charge of data storage	DU
Standards	

Info about metadata	tbd
Standards, Format	WAV for audio, MP4 for video, CSV/TXT for other data sources
Data exploitation and sharing	
Data exploitation	The relation between musical sound and structure and individual bodily movement and biosignals
Data sharing, re-use and distribution	Participants give informed consent to share their data. For any data involving musical performance by professional or otherwise identifiable musicians, sharing will be strictly for research use only.
Archiving and preservation (including storage and backup)	
Data storage (including backup)	Data will be stored on DU servers specifically designed to store and protect this kind of data. Primary storage will be the Music Department's EVO server; backup will be within DU's secure data centre.

Are you generating the data or sourcing it from somewhere else under certain terms and conditions? Data is generated at DU
Is the data digital or non-digital, or both? Digital
How will the data be created or collected? What instruments or tools will be used to produce the data? Audio data recorded using conventional equipment (microphones connected to audio workstations with Avid Pro Tools software, or portable audio multitrack recorders). Video data will be recorded using digital video cameras. Biosignals will be collected using MindMedia NeXus-10 MKII physiological signal recording units. Motion capture data will be collected using the DU Psychology Dept's systems (optical motion capture or body suits). Data will be synchronised using various methods, including timecode, slate-type signals, and custom audio sync tracks.
What transformations will the data undergo? What software or file formats will you use as you work with the data? Data will be edited and synchronised using audiovisual editing software (e.g. Avid ProTools and Media Composer, Apple Final Cut Pro) Data will be analysed using various custom written analytical tools.

Will the data be updated or become redundant as you make revisions and produce subsequent versions? Data will not become redundant, once recorded the data cannot be updated.
Is the data sensitive or confidential? Extracts of musical performances by identifiable (especially professional) musicians may be sensitive commercially or in relation to the musicians' professional reputation. Wherever possible data will be anonymised, but where this is not possible because musicians are identifiable on video images we will ensure that the material is not circulated outside the research team without the musicians' explicit consent.
Is there ethics approval, or is ethics approval required? Ethics approval for the collection of the data has been received in principle from local Research Ethics Committee. Approval for detailed data collection plans will be requested when plans have been finalised.

Finally, for WP3 & WP4 there are further data set from EuroMov as depicted in the table below along with the adequate information.

WP3 Data Acquisition and Multi-Time Signal Analysis and Processing **WP4** Use Case Scenarios	
Tasks **3.1, 3.2, 4.4**	
Data Identification	
Dataset description	Human posture and movements, along with physiological data
Source	Human participants
Partners activities and responsibilities	
Partner owner dataset	EuroMov
Partner in charge of data collection	EuroMov
Partner in charge of data analysis	EuroMov
Partner in charge of data storage	EuroMov
Standards	
Info about metadata	tbd
Standards, Format	CSV or C3D

Data exploitation and sharing	
Data exploitation	We address the relation between movements in solo, duo or group and at multiple time scales
Data sharing, re-use and distribution	Any dataset can be shared on demand, for explicit research purpose. Once anonymized, part of the data might be widely distributed, for example on OpenData servers (case by case analysis is necessary).
Archiving and preservation (including storage and backup)	
Data storage (including backup)	Storage is on EuroMov dedicated and secured computers, with backup on the EuroMov server.

Are you generating the data or sourcing it from somewhere else under certain terms and conditions? EuroMov generates data
Is the data digital or non-digital, or both? Digital
How will the data be created or collected? What instruments or tools will be used to produce the data? We measure movement with VICON mocap, eventually associated to other physiological data (ECG, EMG, EEG, gazing, breathing, and specific data acquisition tools as needed).
What transformations will the data undergo? What software or file formats will you use as you work with the data? We analyse the data using custom written software.
Will the data be updated or become redundant as you make revisions and produce subsequent versions? Original raw data cannot be modified. The data processing pipeline (data curation, pre-processing) includes partly redundant datasets.
Is the data sensitive or confidential? Personal data is always sensitive and confidential. Once anonymized, pre-processed data might no longer be confidential (case by case analysis is necessary).
Is there ethics approval, or is ethics approval required? EuroMov always obtain ethics approval before recording human data for research purposes.

7.1.4 WP5

WP5 Dissemination, Communication and Exploitation	
Data Identification	
Dataset description	Data generated will be digital for Website, Innovation Management Strategy, Plans for Exploitation and Dissemination of results, DMP Questionnaire, Deliverables and Publications (hard copies may also be sent for journal & conference publications), and hard copies in the case of EnTimeMent leaflet and poster.
Source	Data generated during project duration, data from other deliverables, data collected form partners, research data generated, data taken also form the GA
Partners activities and responsibilities	
Partner owner dataset	According to the ownership model.
Partner in charge of data collection	According to the ownership model.
Partner in charge of data analysis	According to the ownership model.
Partner in charge of data storage	According to the ownership model.
Standards	
Info about metadata	TBD
Standards, Format	Format of data generated will be .xls, .ppt, .pdf, .doc and .cdr files and emails.
Data exploitation and sharing	
Data exploitation	TBD
Data sharing, re-use and distribution	Publications, posters or leaflets for dissemination and communication activities based on dissemination rules.
Archiving and preservation (including storage and backup)	

Data storage (including backup)	Data will be stored on encrypted hard disks from each partner, as well in Nextcloud folders

7.1.5 WP6

WP6 Management	WP6 data sets will be further defined and amended later in to the project.