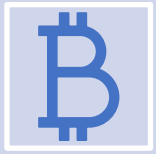
# Data Management Plan

Petia Radeva
Universitat de Barcelona (UB-Tech)

# Index

Motivation and Vocabulary

Data summary: types, format, size, description, origin, metadata, distribution



Data management: strategies, publications, GDPR, Ownership, FAIR, interoperability, accessibility, increase data re-use, data security

# Motivation

**Why do I need to do Data Management Plan?**


A good data management plan will ensure:

- the **availability** and **accessibility** of your results after your project is complete,
- i**ncreasing the value** of your work and eventually
- **possible reuse** by other researchers.

# Motivation

**Definition:** A data and innovation management plan (DIMP), is a formal document that outlines how data will be handled during and after a research project.

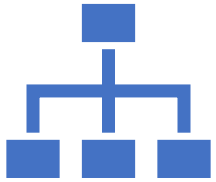
- Many funding agencies, especially government sources, require a DMP as part of their application processes.

**Why is planning important for data management?**

- A DIMP helps achieve **optimal handling, organising, documenting and enhancing of research data**.
  - particularly important for facilitating **data sharing**,
  - ensuring the **sustainability and accessibility** of data in the long-term and
  - allowing **data to be reused** for future research.

- DIMP **protects** the work from **data losses, thefts, and breaches with authentication and encryption tools.**

# Vocabulary



**Data management** provides information on the results of the project, spreading the word about the project successes and outcomes as far as possible.

Dissemination occurs when results and initiatives of the project become available.

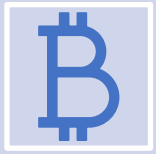
**Intellectual Property Right** refers to the legal rights given to the inventor or creator to protect his invention or creation for a certain period.

These legal rights confer an exclusive right to the inventor/creator or his assignee to fully utilize his invention/creation for a given period.

# Index

Motivation and Vocabulary

**Data summary**: types, format, size, description, origin, metadata, distribution



Data management: strategies, publications, GDPR, Ownership, FAIR, interoperability, accessibility, increase data re-use, data security

# Data summary

- MUSAE claims to empower citizens to improve their health life quality by challenging art and technology symbiosis elaborating innovative business models, technologies and guidelines of use.

- This involves:
  - the design and i**mplementing of proposed solutions i**n the selected proposals during the Open Calls of MUSAE to test their effect, as well as
  - gathering **input and feedback from citizens and users** to improve solutions, increase **user acceptance, encourage behavioural change,** and achieve **healthier lifestyle** and/or i**mproved sustainable food solutions** in the market.

# Data types, formats and size

- **Manually collected data:** demographic data, data on nutritional lifestyle, pictures, audio and video done by citizens, etc.
  - Partner representatives
  - External individuals who voluntarily participate in a residency and pilot activities.

- **Data automatically collected through technology:** AI, wearables and robotics.

Data should be organised in datasets relating to the category of the data and site of collection.

**Only data that is needed** to perform project activities will be collected, and as far as possible, participants will not be asked to provide personal data unless this is necessary.

# Description of data

Part of these data are to be **open and shared** through the Factory model pack and the related guidelines.

## Data formats

A dataset can include different types of formats e.g., a manually collected dataset concerning user acceptance can consist of both written interview notes, audio files from interviews, pictures from art-tech experiments, and survey responses.

- Some of this data **cannot be anonymised** within the scope of this project (e.g. audio files), so in most cases only parts of a dataset can be made openly available.

MUSAE will only **use widely accepted formats** for data generation, such as:

- Documents/Reports/Publications: .PDF/A, txt, doc/docx
- Spreadsheets: .xls/.xslx
- Databases: .cvs
- Audio files. .mp3, .wav, .wma, .ra
- Pictures: jpg, png, bmp, tiff
- Video: avi, flv, mov, mp4, wmv

# Origin of data

Each proposal by MUSAE will collect its necessary data according to its main goal and target. Depending on the type of data, there will be various methods and origins of data collection involved at each art-tech experiment.

- For **manually collected data,** the main origins could be:
  - - Interviews with groups and individual participants in the selected proposals, if relevant for the corresponding proposal
  - - Survey responses, if relevant for the corresponding proposal
  - - Literature study/review and open data (re-use of existing data), etc.

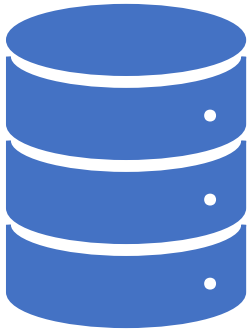
- For **automatically collected data** the main origins could be:
  - - Automated data collection by one of the technologies: AI, robotics and wearables managed by the corresponding pilot.
  - - Mobile phone applications voluntarily downloaded by participants, etc.

- The non-anonymous datasets should be **stored locally by the Data Controllers** of the selected proposals and **not shared with others,** except for project generated contact lists which will be stored in a strict access controlled MUSAE repository.

# Metadata of a pilot-used dataset

| DATA CHARACTERISTIC | DESCRIPTION |
| --- | --- |
| File/Dataset name | |
| Dataset origin | |
| Dataset purpose | |
| Version | |
| Description | |
| Pilot | |
| Date | |
| File type | |
| Pilot leader(s) | |
| Data controller | |
| Data size | |
| Data sharing | |
| Archiving and preservation | |
| Re-used existing data | |
| Data utility | |
| Link to the data | |

# Data distribution



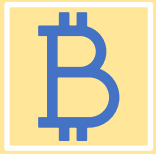
The data sharing should contain an explanation of the sharing policies related to the dataset

Each **dataset must have its distribution license**.

# Index

Motivation and Vocabulary

Data summary: types, format, size, description, origin, metadata, distribution



**Data managemen**t: strategies, publications, GDPR, Ownership, FAIR, interoperability, accessibility, increase data re-use, data security

# Data and Knowledge Management Strategies

- Data storage
- Data collected in the project will come from three main sources:
  - MUSAE selected proposals,
  - public datasets and
  - possible external data providers.
- Annotations, curation and de-identification will be performed locally by the MUSAE selected proposals with open access or home-made software.

- The MUSAE consortium strongly encourages pilot leaders and data controllers to use standard open-access platforms **like Zenodo (https://zenodo.org), for data storage so that de-identification tasks can be carried-out using an Anonymizer.**

# Data and Knowledge Management Strategies

Open access to research data

The final outcome of a STARTS open call shall be made **available at no cost to the EC directly and for public viewing in exhibitions.**

The **EC has the right to show them for at least a year after the end of the residency**.

# Open access to peer-reviewed scientific publications

MUSAE research partners publish scientific publications including project results in Open Access.

- Open access can be defined as the practice of providing **on-line access** to scientific information that is free of charge to the end-user.

All scientific publications generated by the MUSAE project will be made available online through open access in peer reviewed scientific journals.

# GDPR Compliance Principles

- The project will be fully compliant with the European Union (EU) 2016/679 General Data Protection Regulation (GDPR) and AI-Act ensuring:

- *Lawfulness, fairness and transparency*:

- *Purpose limitation*

- *Data minimization*

- *Accuracy, storage limitation, integrity and confidentiality*

- *Accountability*

# Data ownership

Any data gathered during the lifetime of the selected proposals are under the **ownership of the beneficiary or the beneficiaries** that produce them.

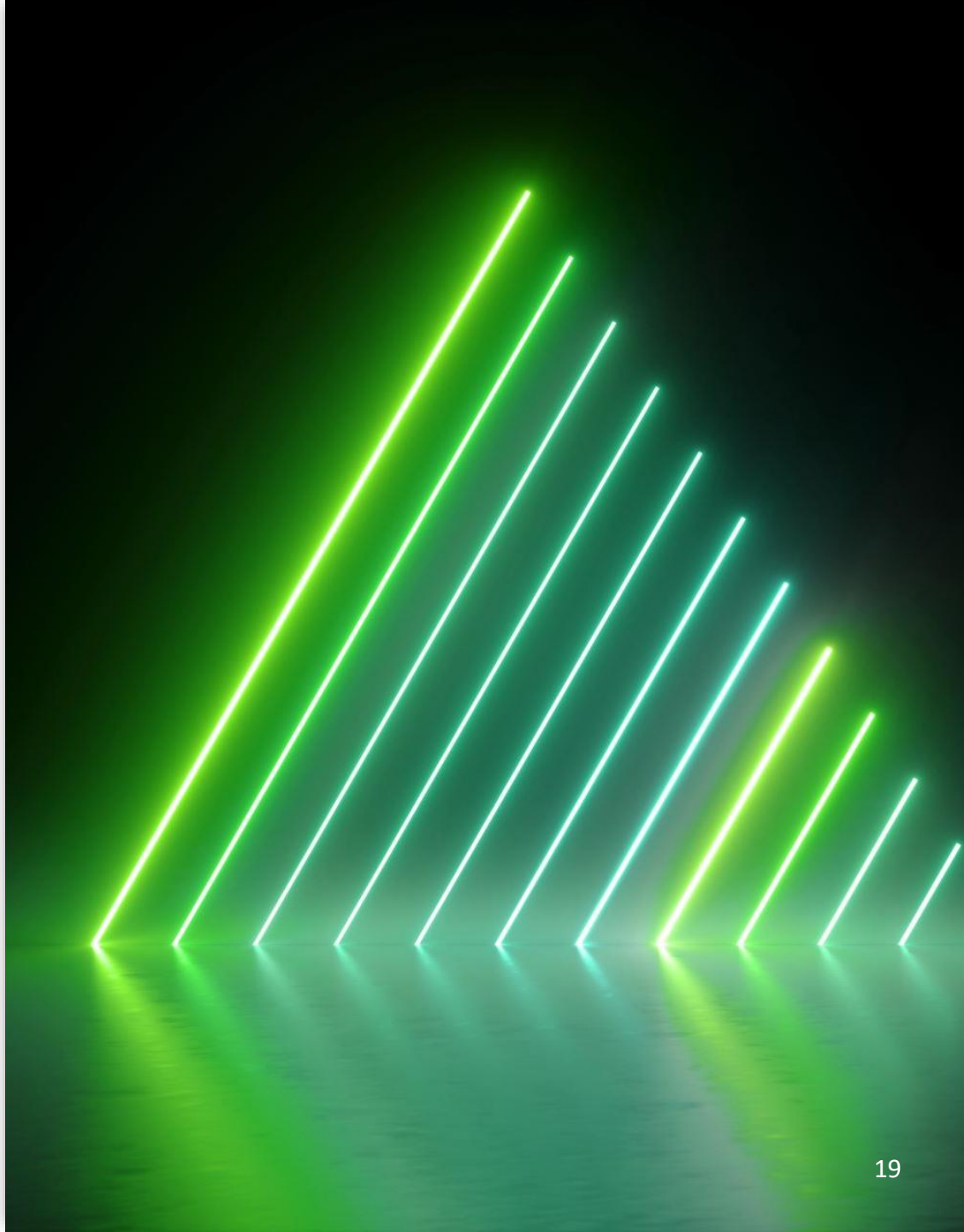
The beneficiaries have the **intellectual property rights** of the data they collect and re-use of data is defined by the limitation they might set in how they will make data available.

This means **partners decide if they make their data freely and openly available** to the research community or there is an embargo period, whereby permission for accessing the data is given after a certain period.

# FAIR Data management

MUSAE assures a FAIR approach to all collected data:

- **Findable**
- **Accessible**
- **Interoperable**
- **Reusable**

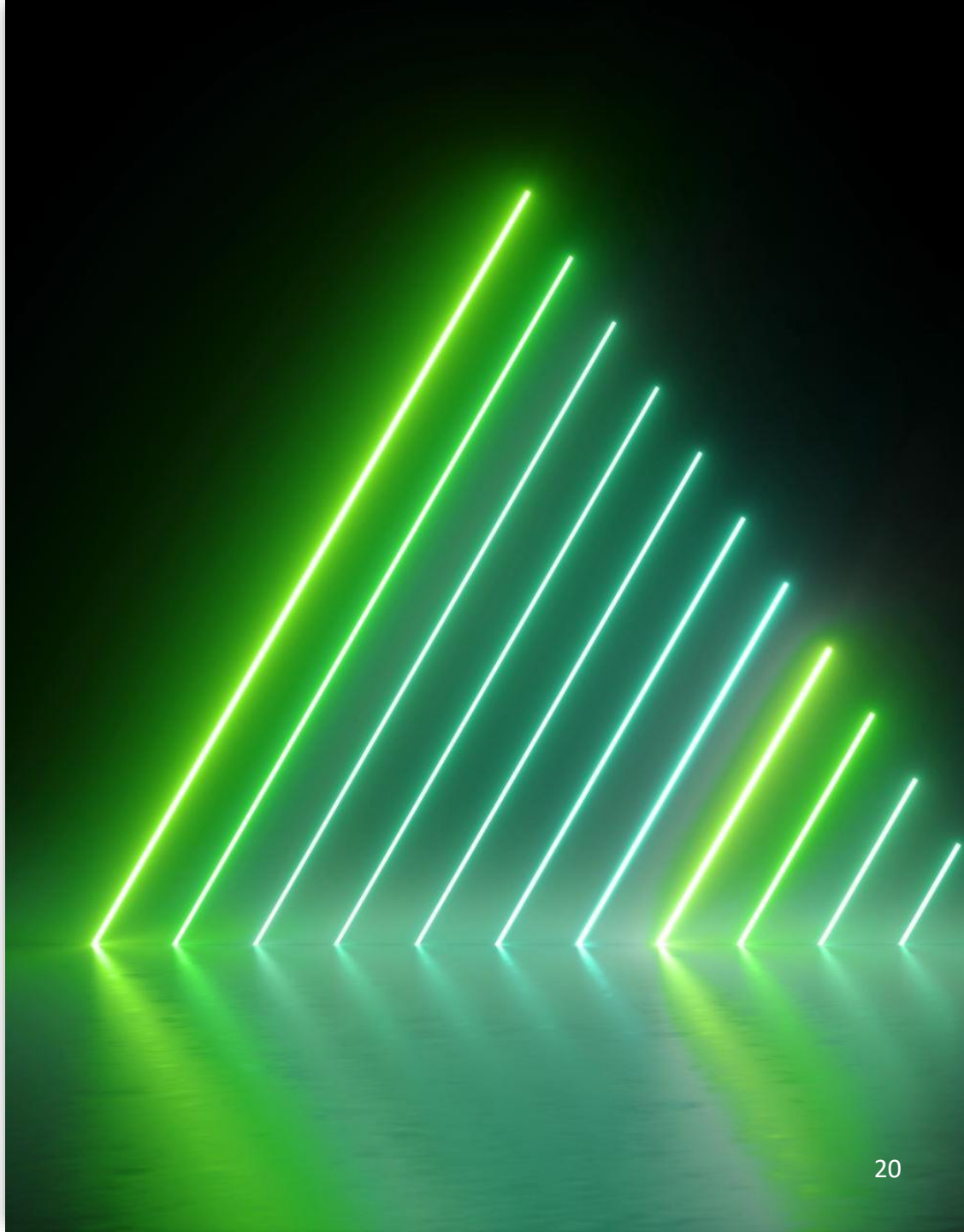


# FAIR Data management

MUSAE assures a FAIR approach to all collected data:

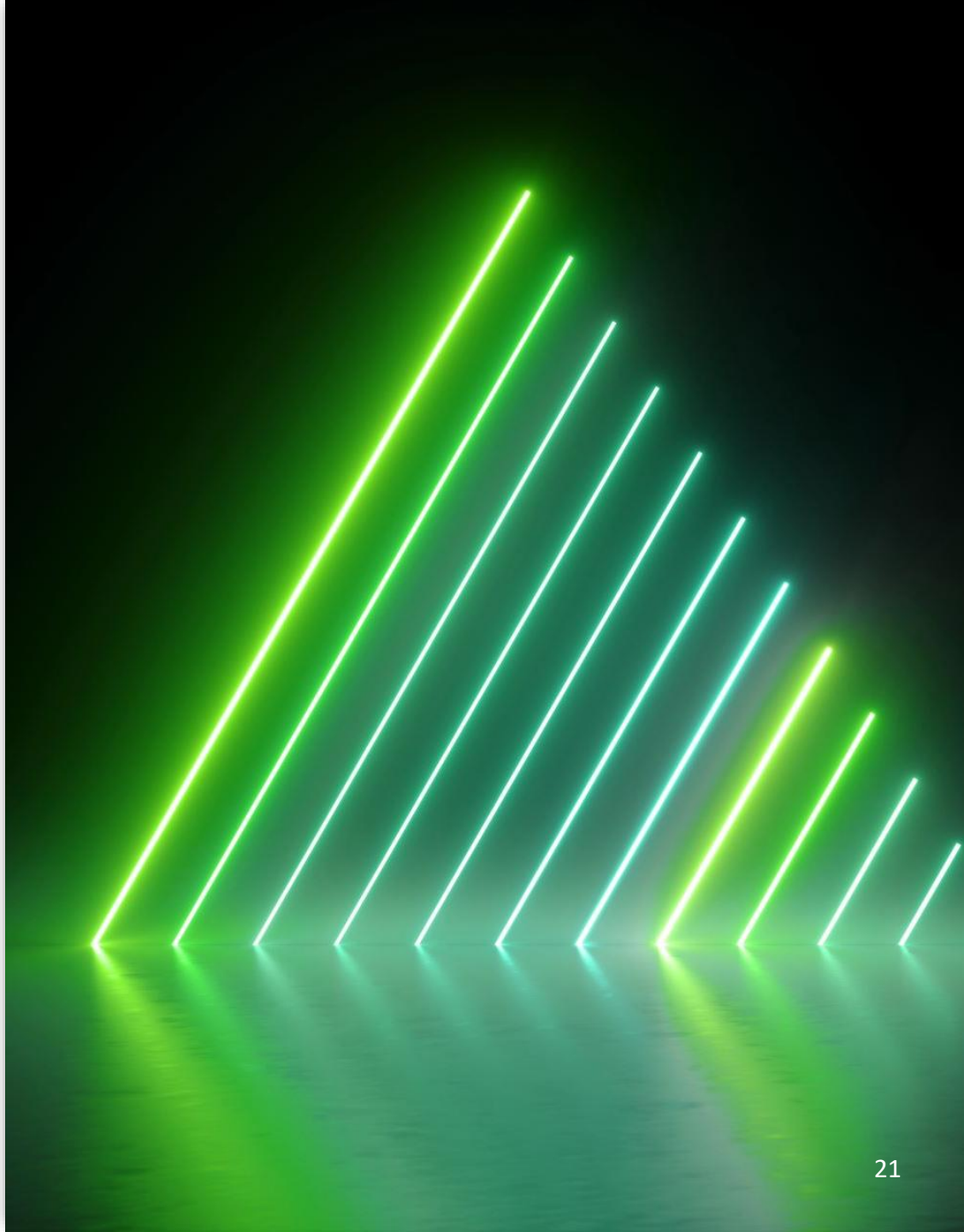
- **Findable** (stimulating data discovery): all selected proposals' partners should agree to provide access to data generated within the project (apart confidential data protected by IPRs, as introduced above).

- **Accessible** (managing access of sensitive data): The MUSAE project will ensure that *shareable* data are accessible, but when data will contain sensitive information, the access must be controlled.



# FAIR Data management

- MUSAE assures a FAIR approach to all collected data:
  - **Interoperable:** MUSAE is expected to gather *highly heterogeneous* data due to the variety of the selected proposals, solutions and technologies expected to be developed within the project.
  - **Reusable:** Open data from the MUSAE experimentation is to be made *publicly available* through the (E)DIHs network and the STARTS centres and *reusable to transfer* and apply the DFA model.



# Metadata in Zenodo

To make data findable, the main tool is to assure that data used and produced in the project can be discoverable with **metadata.**

# Approach to search keywords

**Data Controllers** at each project will be responsible for uploading public datasets that they have generated and to assign specific keywords relevant to these datasets to make them searchable.

Datasets will be named using the following naming conventions:

*DS_PilotCode_DataCategoryNr_DataController_Description_Acronym_UniqueDataNr*

# Data security

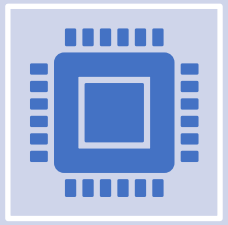
MUSAE will provide out-of-the-box security mechanisms and management procedures so as to

a) **ensure personal (sensitive) data protection** through a strict process of data collection, anonymization, harmonization and integration and

b) **guarantee data integrity and reliability.**

# Data security

Access will always be **restricted to the personnel involved** in the study using encrypted passwords.



Moreover, the **user access will depend on the roles** within the project.



**Metadata and data stored** in specialised repositories should be **duplicated and backed up** and provide **secure means of data transfer.**

# Conclusions

- As responsible for the DIMP of MUSAE, **UB will ensure** that **any data management** issues which may arise during the project will be **handled appropriately** and in a transparent and fair manner.

- The **DIMP is a living document** that will expand as the project evolves and new information on data collection, generation and handling arise.

- Day to day, data management will happen through c**ontinuous collaboration between the coordinator** (POLIMI), the proposals **Data Controllers**, the **ethical task leader** (ABACUS), and the **DIMP leader** (UB-Tech).

# Key MUSAE contacts

| Participant No | Organisation Name | Main Contact | Email | Alternative contact | Email |
|---|---|---|---|---|---|
| 1 | POLIMI | Maria Rita Canina | marita.canina@polimi.it | Tatiana Efremenko<br>Carmen Bruno | tatiana.efremenko@polimi.it<br>carmen.bruno@polimi.it |
| 2 | ABACUS | Maria Bulgheroni | mariabulgheroni@ab-acus.com | Laura Giani | lauragiani@ab-acus.eu |
| 3 | UoM | Angelo Cangelosi | angelo.cangelosi@manchester.ac.uk | Marta Romeo | marta.romeo@manchester.ac.uk |
| 4 | UB | Petia Radeva & Eloi Puig | petia.ivanova@ub.edu<br>eloi35@gmail.com | Teresa Minguell &<br>Umair Haroon | teresawxyz@gmail.com<br>umairharoon3797@gmail.com |
| 6 | PAL | Gizem Bozdemir | gizem.bozdemir@pal-robotics.com | Lorna McKinley | lorna.mckinlay@pal-robotics.com |
| 7 | GLUON | Ramona Van Gansbeke | ramona@gluon.be | Christophe De Jaeger | christophe@gluon.be |
| 8 | UCD | Prof. Lorraine Brennan | lorraine.brennan@ucd.ie | Aoife O' Gorman | aoife.ogorman@ucd.ie |

# THANK YOU

MUSAE

European Commission