conbots
CONnected through roBOTS

Deliverable D1.1

link to the on-line version that undergoes ongoing modifications is provided here
(password: conbotsdmp)

Data Management Plan

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N. 871803

WP1 - Project management and coordination

TASK 1.1 - General Management of the project

D1.1 - Data Management Plan (DMP)

Contract number:	871803
Project acronym:	CONBOTS
Project title:	CONnected through roBOTS: physically coupling humans to boost handwriting and music learning
Planned delivery date:	July 1st, 2020
Author(s):	CONBOTS consortium
Partners contributed:	ALL
Document date:	June 29th, 2021[1]
Version:	3
Revision:	0
Deliverable type:	Report
Abstract:	Definition of the Data Management Plan (DMP), as foreseen by the "Pilot on Open Research Data" of Horizon 2020. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
Status:	☒ PU (Public) ☐ PP Restricted to other programme participants (including the Commission Services) ☐ Restricted to a group specified by the consortium (including the Commission Services) (please specify the group) ☐ Confidential, only for members of the consortium (including the Commission Services)

[1] Due to the COVID outbreak, most of project activities have been slowed down by the extensive lockdown period and have been significantly delayed. We contacted the PO asking for an extension of all the project activities and deadlines, including the submission of the deliverables foreseen at M3, M4 and M6.

1. DOCUMENT REVISION LOG

VERSION	REVISION	DATE	DESCRIPTION	AUTHOR
0.1	0	19/10/2020	First Draft: structure of plan according to template.	IBM
0.2	1	2/2/2021	Incorporating additions of data by GU	IBM
0.3	2	8/2/2021	Incorporating additions of data by IUVO	IBM
0.4	3	11/2/2021	Incorporating additions of data by UCBM	IBM
0.5	4	15/2/2021	Incorporating additions of data by SSSA	IBM
1	5	22/2/2021	Version sent for internal consortium review prior to submission	IBM
1.1	6	7/3/2021	Incorporating additions by ARVRtech	IBM
1.2	7	8/3/2021	Submission-ready	IBM
2	0	22/3/2021	General revision	UCBM
2	1	21/6/2021	Comments and requests for finalization	UCBM
2	2	23/6/2021	Revisited and made available on-line	IBM
3	0	29/06/2021	Version ready for submission with live document available online	UCBM

2. CONTENTS

3. LIST OF ACRONYMS AND ABBREVIATIONS

DMP – Data Management Plan.

FAIR – findable, accessible, interoperable, and re-usable.

4. INTRODUCTION

This report is the CONBOTS Data Management Plan (DMP). This DMP describes the data management life cycle for the data to be collected, processed and/or generated by CONBOTS. As part of making its research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on the handling of research data during & after the end of the project. This includes elaborating about what data will be collected, processed and/or generated, which methodology & standards will be applied, whether data will be shared/made open access and how data will be curated & preserved (including after the end of the project).

Overall, the DMP report is meant to be a description of the handling of data during and subsequent to project completion, of the data that is collected, generated, and processed, curated, and persisted. Related method and standards applied will also be reported where relevant.

The structure of the document conforms to the EC template for a DMP as can be found here:

Guidelines on FAIR Data Management in Horizon 2020 are available in the Online Manual.

As suggested, this document is gradually evolved to provide answers to the set of questions in the template. The DMP is as such a living document in which information is made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. We keep track of version numbers upon its time of updating, and ensure it is also revisited prior to the periodic evaluation/assessment of the project, and also before its final review.

A link to the on-line version that undergoes ongoing modifications is provided here (password: conbotsdmp).

5. Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

5.1. Data collection purpose and relation to objectives

What types and formats of data will the project generate/collect?

Will you re-use any existing data and how?

What is the origin of the data?

What is the expected size of the data?

To whom might it be useful ('data utility')?

Overall, the purpose of data collection in CONBOTS is threefold. First, data captured is meant to be a basis for developing machine learning models in order to identify the users' physical, mental, and emotional states while in a learning process. Second, data is collected to build models that could facilitate the working of the interaction platform (i.e., exoskeleton) to improve learning effectiveness, while ensuring aspects such as data encryption and privacy of subjects' information. The development of the latter models for interaction will attempt to generate data to make it pragmatic to build on previous knowledge in both game-theory and deep reinforcement learning. Finally, data will be collected from the experiments for the validation of the CONBOTS platform, and it will be used to assess the feasibility of using the CONBOTS platform as a tool for supporting handwriting and music learning.

In relation to project objectives, CONBOTS is pursuing a variety of experiments to gradually capture the data needed. This yields different types of data being collected. Specifically, data collection includes the following elements:

Element ID: 1
Origin: Recording of music scenarios by Ghent University
Type(s): Audio, video (e.g., of the participants in music scenarios), kinematic data (120hz, 42 markers on MoCap suit), Questionnaires (open and Likert scale answers), eye tracking data.
Formats: • Comma Separated Values (.csv) • Matlab files (.mat) • R files (.r) • Excel files (.xlsx) • FlimBox (.fbx)
Element ID: 2
Origin: Recording of data by IUVO
Type(s): 1. Signed informed consent templates 2. Pseudonymized personal data (e.g., age, gender of participants) 3. Pseudonymized anthropometric data (e.g., height, weight, shoulder breadth, arm length) 4. Pseudonymized kinematic data (e.g., shoulder range of motion, angular velocity and acceleration)

5. Pseudonymized score or open answers as output of ad hoc questionnaires (e.g., on usability, wearability of the exoskeleton)
6. Pseudonymized picture/video of participants during experimental activities
7. Pseudonymized exoskeleton wearing data (e.g., regulations set)
8. Pseudonymized exoskeleton mechatronics data:
    - data from joint encoders (e.g., position and related data like joint velocity and acceleration)
    - joint torques
    - in case of exoskeleton active module (e.g., shoulder flex/extension): actuators current, voltage and parameters related to the control system

Formats:

1. Anonymized picture/video (e.g., jpg, tiff, png, avi, mp4)
2. MATLAB file (Mathworks, Natick, MA, USA; e.g., mat, csv)
3. Excel file
4. Exoskeleton-related data will be generated by means of Labview 2014 or Labview 2018 (National Instrument, Austin, TX, USA) as .bin binary files, that can be read in MATLAB environment.

Element ID: 3

Origin: Recording of data by UCBM

Type(s):

- ECG, respiratory signal, GSR, M-IMU data (free accelerations, angular speed, quaternions); sampling rate: 100 Hz;
- EMG (sampling rate: 1.1 kHz);
- Robot's Kinematic and Dynamic Data (position, orientation and interaction forces; sampling rate: 100 Hz);
- User performance data (video-game cursor and target kinematic data, position and orientation errors); (sampling rate: 100 Hz)
- Position of the pen on graphics tablet during handwriting exercises
- Standard tests for the assessment of handwriting (DGM-P, BHK)
- Anthropometric data (e.g., height, weight, shoulder breadth, arm length)
- Pictures, audio and video of the experiments
- Score or open answers as output of ad hoc or standardized questionnaires (e.g., task difficulty, perceived effort, emotional expressivity, GOLD-MSI index, comfort rating scale, etc.)
- Signed informed consent templates

Formats:

1. Pseudonymised data in .csv , .mat , .txt and .xlsx, .mp4, .wav and .png format.

Element ID: 4

Origin: Recording of data by SSSA

Type(s):

We will collect kinematic (i.e. joint angles, velocities,…) and kinetic (i.e. torques) data from the elbow module of the exoskeleton. Additionally, during experiments, we may also want to collect IMU and EMG data, and also pictures and videos.

Formats:

In general, we save our data in .bin files. However, they can be easily converted in .csv format for sharing purposes.

Element ID: 5

Origin: Recording of data by ARVRtech

Type(s): • Pseudonymized Personal data (e.g., age, gender of participants) • Pseudonymized Anthropometric data (e.g., height, weight, shoulder width, arm length) • Pseudonymized picture/video of participants during experimental activities • Motion Capture files (transformation of user joints in time)
Formats: • .json (user data, motion data from motion capture session)

We further elaborate additional info related to each data element in the following sections.

6. FAIR data

6.1. Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

What naming conventions do you follow?

Will search keywords be provided that optimize possibilities for re-use?

Do you provide clear version numbers?

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

We number the various data elements in this report and indicate its origin. As we proceed with a more concrete data acquisition, individual data elements may be further associated with version numbers. Correspondingly, concrete data storage components (e.g., file names, database names, spreadsheets) will be identified with a naming convention that associates it with its creator and date of creation. Relevant meta-data and search keyword will be included here to simplify the use and finding of the data.

6.2. Making data openly accessible

This section details data elements that are made openly available, and if cannot be shared, clarify legal, contractual, or other restrictions the prohibit it from being shared[2].

For each data element, we further elaborate where it is deposited, methods or software tools needed to access the data, restrictions on use, conditions for access, etc.

Element ID: 1
Origin: Recording of music scenarios (Ghent University)
List of all related parts of this data elements that will be made openly accessible and via which repository (e.g., Zenodo): Until now we shared our data with ARVRtech lab through Belnet transfer as advised by UGhent data stewards. We do not have experience with Zenodo, OpenAIRE or CERN platforms.

[2] Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

<<To be elaborated in subsequent versions of this doc, including restriction on use and conditions for access i.e., machine readable license>>
List of all related parts of this data elements that cannot be shared and corresponding clarifications. <<To be elaborated in subsequent versions of this doc>> There is no foreseen intention to share this data element / parts of it with someone external to CONBOTS. If any data is deemed relevant for sharing, only processed, fully anonymized data will be included e.g., summary statistics.
Method / software tools required to access the data: *All data is being processed and analysed in Matlab, R, Qualysis, Ableton etc.* Corresponding software documentation to access that data: <<…>>, links to open source: <<…>>
Element ID: 2
Origin: Recording of data by IUVO
List of all related parts of this data elements that will be made openly accessible and via which repository (e.g., Zenodo): At the moment making data public is not foreseen. In the future, in case of results dissemination (e.g., by means of scientific papers) it could be possible to make experimental data public. For instance, papers could be published (i) about mechatronic-related results of the exoskeleton, and/or (ii) involving healthy subjects testing the exoskeleton. In the latter case, participants will be aware about it and they will be asked to sign a consent form before being involved in the experimental activities; experimental procedure will be approved by the Local Ethical Committee. <<To be elaborated in subsequent versions of this doc, including restriction on use and conditions for access i.e., machine readable license>>
List of all related parts of this data elements that cannot be shared and corresponding clarifications. <<To be elaborated in subsequent versions of this doc>> At the moment sharing data with entities external to CONBOTS is not foreseen. In the future, in case of results exploitation it could be possible to share anonymous data/parts with third parties (e.g., Ossur and Comau as IUVO shareholders and possibly interested in commercializing solutions coming from the project).
Method / software tools required to access the data: *All data is being processed and analysed in Matlab, R, Qualysis, Ableton etc.* Corresponding software documentation to access that data: <<…>>, links to open source: <<…>>
Element ID: 3
Origin: Recording of data by UCBM
List of all related parts of this data elements that will be made openly accessible and via which repository (e.g., Zenodo): Part of the collected data will be shared (pseudonymised) with collaborators external to the consortium, employed at the Italian National Research Council (CNR) and University of Milan. Only processed, fully anonymized data (e.g., summary statistics) will be made public. Pseudonymized data are stored in hosting platforms available for UCBM staff (Dropbox and Onedrive) and protected by the personal institutional accounts of the project coordinator and other researchers involved in

the studies. For data sharing, Data will be protected using AES 256 encryption and then shared with platforms implementing end-to-end encryption, such as Tresorit. <<To be elaborated in subsequent versions of this doc, including restriction on use and conditions for access i.e., machine readable license>>
List of all related parts of this data elements that cannot be shared and corresponding clarifications. <<To be elaborated in subsequent versions of this doc>>
Method / software tools required to access the data: All data is being processed and analyzed in Matlab. Corresponding software documentation to access that data: <<…>>, links to open source: <<…>>
Element ID: 4
Origin: Recording of data by SSSA
List of all related parts of this data elements that will be made openly accessible and via which repository (e.g., Zenodo): Data will not be shared outside the CONBOTS consortium with the exception of the potential publication of results in scientific publications. Results (e.g. aggregated data) from experimentations can be made public, through scientific publication, conferences and, more in general, dissemination events. <<To be elaborated in subsequent versions of this doc, including restriction on use and conditions for access i.e., machine readable license>>
List of all related parts of this data elements that cannot be shared and corresponding clarifications. <<To be elaborated in subsequent versions of this doc>>
Method / software tools required to access the data: Corresponding software documentation to access that data: <<…>>, links to open source: <<…>>
Element ID: 5
Origin: Recording of data byARVRtech
List of all related parts of this data elements that will be made openly accessible and via which repository (e.g., Zenodo): At the moment making data public is not foreseen. In the future, in case of results dissemination (e.g., by means of scientific papers) it could be possible to make experimental data public. <<To be elaborated in subsequent versions of this doc, including restriction on use and conditions for access i.e., machine readable license>>
List of all related parts of this data elements that cannot be shared: End-user personal data will not be shared with any third party. Motion capture files might be shared with the partners on the project. (Ghent, etc.) <<To be elaborated in subsequent versions of this doc>>
Method / software tools required to access the data: Corresponding software documentation to access that data: <<…>>, links to open source: <<…>>

6.3. Making data interoperable

For any piece of data that may be exchanged or re-used between researchers, institutions, organizations, countries, etc, we hereby elaborate corresponding information. This includes data exchange formats, compliance to relevant (open) software applications etc. This may also include description of metadata vocabularies, standard or methods to make data interoperable. When using non-standard vocabularies, we also provide links to more commonly used ontologies.

Element ID: 1
Origin: Recording of music scenarios (Ghent University)
Data exchange formats: for descriptive data .csv, for MoCap data .mat, .fbx.
Peers with whom data should be shared with: all partners that develop technology for 2 music scenarios e.g., ArVrtech, IUVO, SSSA.
References to meta data / schemas / ontologies: <<To be elaborated in subsequent versions of this doc>>
Methods applicable to data handling: adhering to the guidelines of Ghent University that are based on EU's GDPR.
Element ID: 2
Origin: Recording of data by IUVO
Data exchange formats: 1. Anonymized picture/video (e.g., jpg, tiff, png, avi, mp4) 2. MATLAB file (Mathworks, Natick, MA, USA; e.g., mat, csv) 3. Excel file 4. Exoskeleton-related data will be generated by means of Labview 2014 or Labview 2018 (National Instrument, Austin, TX, USA) as .bin binary files, that can be read in MATLAB environment
Peers with whom data should be shared with: 1. SSSA (pseudonymized) 2. UCBM (pseudonymized) 3. GU (pseudonymized) 4. IBM (pseudonymized)
References to meta data / schemas / ontologies: <<To be elaborated in subsequent versions of this doc>>
Methods applicable to data handling: Personal data will be pseudonymized by the authorized by the research staff. Then, they will be stored in encrypted folders, whose access is limited to authorized staff by individual passwords. A data sharing agreement will be signed within the protocol submission before the competent Ethics Committee.
Element ID: 3
Origin: Recording of data by UCBM
Data exchange formats: We use the following file format: .csv , .mat, .txt, .wav, .mp4., png.
Peers with whom data should be shared with:

We will share our data with all the partners that need to use it for the project tasks in which they are involved, and in particular with IBM, ICL, IUVO, SSSA and GU.
References to meta data / schemas / ontologies: <<To be elaborated in subsequent versions of this doc>>
Methods applicable to data handling: We are following the guidelines of the Università Campus Bio-Medico di Roma (UCBM) that are based on EU's GDPR. Moreover, UCBM appointed a DPO for the project and an external Data Protection Advisor, that indicates and controls all the relevant procedures needed for data management and sharing.
Element ID: 4
Origin: Recording of data by SSSA
Data exchange formats: In general, we save our data in .bin files. However, we believe that .csv formats could be reasonable.
Peers with whom data should be shared with: Probably, the data coming from the exoskeleton could be shared with partners that have to interact with the exoskeletal platform, e.g. IUVO, UCBM, GU, ICL, IBM.
References to meta data / schemas / ontologies: <<To be elaborated in subsequent versions of this doc>>
Methods applicable to data handling: Referring to the processing of exoskeleton/biomechanical data collected during experiments, we typically process them with custom Matlab routines. Otherwise, referring to the handling of personal data, we are only aware of general guidelines for pseudo-anonymization.
Element ID: 5
Origin: Recording of data by ARVRtech
Data exchange formats: We save our data in .json files on the server.
Peers with whom data should be shared with: Motion capture data could be shared with partners, e.g. IUVO, UCBM, GU, ICL, IBM.
References to meta data / schemas / ontologies: <<To be elaborated in subsequent versions of this doc>>
Methods applicable to data handling: Motion capture and error calculation data collected by us is a custom model of data written by us. We could export motion data to some standard motion data type (e.g. .fbx).

6.4. Increase data re-use (through clarifying licences)

In this section we relate to the enablement of data re-use. This includes possible use by third parties, licensing that permits its re-use, planned scheduling for its availability and archiving. If an embargo is sought to give time to publish or seek patterns. We also specify why and how long it applies. Corresponding data quality assurance processes may also be described when relevant.

Element ID: 1
Origin: Recording of music scenarios (Ghent University)
Data release / archiving: All data generated during the CONBOTS project will be archived and stored on the password protected IPEM server. Storage is guaranteed for 5 years after the end of the research (UGhent standard policy). Additionally,

all the data gathered in ASIL laboratory is directly upload on a storage with a capacity of 27,47 TB. The offsite backup is physically located in the Faculty of Engineering in Zwijnaarde (University of Ghent). The data will be discarded 5 years after the end of the CONBOTS project.
Element ID: 2
Origin: Recording of data by IUVO
Data release / archiving: Consent form will be archived for 5 years, then they will be discarded. Experimental data will be archived pseudoanonymized for the project duration, then they will be archived anonymized (i.e., data subject is no longer identifiable). Consent form will be stored for 5 years, access will be limited to the p.i. and his/her authorized staff. They could be accessible only in case of audits from the competent authorities / participants' rights exercise. After 5 years, they will be destroyed.
Element ID: 3
Origin: Recording of data by UCBM
Data release / archiving: Consent form will be archived for 5 years after the end of the project, then they will be discarded. Experimental data will be archived pseudonymized for the project duration, then they will be archived anonymized (i.e., data subject is no longer identifiable). Consent form will be stored for 5 years, with access limited to the PI of the study and his/her authorized staff. They could be accessible only in case of audits from the competent authorities / participants' rights exercise. 5 years after the end of the project, they will be destroyed.
Element ID: 4
Origin: Recording of data by SSSA
Data release / archiving: Raw pseudo-anonymized data could be stored for 7 years after the completion of a study; informed consents after 5 years. All personally identifiable information could be deleted after 7 years after the end of a study, thus maintaining the raw data in completely anonymized form from then on.
Element ID: 5
Origin: Recording of data by ARVRtech
Data release / archiving: The consent form will be archived for 5 years after the end of the project, then they will be discarded. Experimental data will be archived pseudonymized for the project duration, then they will be archived anonymized (i.e., data subject is no longer identifiable). The consent form will be stored for 5 years, with access limited to the PI of the study and his/her authorized staff. They could be accessible only in case of audits from the competent authorities/participants' rights exercise. 5 years after the end of the project, they will be destroyed.

7. Allocation of resources

What are the costs for making data FAIR in your project?

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

Who will be responsible for data management in your project?

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

<<To be elaborated in subsequent versions of this doc>>

8. Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Is the data safely stored in certified repositories for long term preservation and curation?

<<To be elaborated in subsequent versions of this doc>>

9. Ethical aspects

Sharing of data is overseen by ethics review when deemed necessary. In this section we detail whether any part of the data may include private data that requires such approvals, and provide corresponding references to ethics approvals, informed user consents, and related questionnaires on long term data preservation.

Element ID: 1
Origin: Recording of music scenarios (Ghent University)
Inclusion of personal data: yes
Description of data that may require ethics approval: data may include attributes such as sex, profession, education, background, etc. We will also collect kinematic, video, audio and personal data of children. Therefore, following guidelines of UGhent data stewards, we will share only (pseudo)anonymized data with our partners. Also, we will share these data only with the partners that need it for their work e.g., ArVrtech, IUVO, SSSA.
Related standards, policies, legislation etc.: Data conforms to guidelines of Ghent University that are based on EU's General Data Protection Regulation (GDPR).
Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc: <<To be elaborated in subsequent version of this doc>>
Element ID: 2
Origin: Recording of data by IUVO
Inclusion of personal data: yes
Description of data that may require ethics approval: All data are collected during experiments by IUVO staff committed to specific confidential obligations. Once pseudonymized, research data are shared to other partners.

Part of the activities are related to experimental procedures involving human beings. Participants will be informed about experimental procedures, including personal data collection, processing, and storage. If they consent, they will participate. All details will be included in the protocol to be submitted to the competent ethical committee.
Related standards, policies, legislation etc.: As far as personal data are concerned: Ethics Code on personal data processing for scientific research and statistics purposes issued by the Italian Data Protection Authority, the DLGS 196/2003 on Privacy Code as amended by the DLGS 101/2018, and the EU Reg. 2016/679 General Data Protection Regulation. Specific technical and organizational measures emerged from the data protection impact assessment.
Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc: <<To be elaborated in subsequent version of this doc>>
Element ID: 3
Origin: Recording of data by UCBM
Inclusion of personal data: yes
Description of data that may require ethics approval: We do not collect private data. We collect the following personal data: sex, profession, education, background, handiness and previous upper limb injuries. Part of the activities are related to experimental procedures involving human beings. The experimental protocols are approved by UCBM Ethics committee which assures that all the procedures comply with relevant regulations and possible ethical issues. Participants will be informed about experimental procedures, including personal data collection, processing, and storage. Volunteers will participate in the study only if they read, understand and sign the informed consent approved by the Ethics committee.
Related standards, policies, legislation etc.: Technical standards are the ones associated with the file format mentioned above. As regards data collection, storage and sharing, we comply with legal and ethical aspects on personal data processing for scientific research and statistics purposes issued by the Italian Data Protection Authority, the DLGS 196/2003 on Privacy Code as amended by the DLGS 101/2018, and the EU Reg. 2016/679 General Data Protection Regulation. Moreover, all the experiments involving human subjects are approved by the UCBM ethics committee.
Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc: <<To be elaborated in subsequent version of this doc>>
Element ID: 4
Origin: Recording of data by SSSA
Inclusion of personal data: no
Description of data that may require ethics approval: All the data collected will pertain to exoskeleton users and will not be personally attributable to him/her. We will have pseudo-anonymized data. If the data shared between partners are in an anonymized way, we believe that no particular ethics issue should arise.

Related standards, policies, legislation etc.: We are aware of the GDPR regulation about protection of personal data.
Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc: <<To be elaborated in subsequent version of this doc>>
Element ID: 5
Origin: Recording of data by ARVRtech
Inclusion of personal data: yes
Description of data that may require ethics approval: We will have anonymized data. If the data shared between partners are in an anonymized way, we believe that no particular ethics issue should arise.
Related standards, policies, legislation, etc.: We are aware of the GDPR about the protection of personal data.
Link to related docs - Ethics approval, consents, questionnaires, long-term preservation etc: <<To be elaborated in subsequent version of this doc>>

10. Other issues

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

<<To be elaborated in subsequent versions of this doc>>

11. Further support in developing your DMP

The Research Data Alliance provides a Metadata Standards Directory that can be searched for discipline-specific standards and associated tools.

The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

Useful listings of repositories include:

Registry of Research Data Repositories

Some repositories like Zenodo, an OpenAIRE and CERN collaboration), allow researchers to deposit both publications and data, while providing tools to link them.

Other useful tools include DMP online and platforms for making individual scientific observations available such as ScienceMatters.

<<To be elaborated in subsequent versions of this doc>>

SUMMARY TABLE 1

FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined above.

DMP component	Issues to be addressed
1. Data summary	• State the purpose of the data collection/generation • Explain the relation to the objectives of the project • Specify the types and formats of data generated/collected • Specify if existing data is being re-used (if any) • Specify the origin of the data • State the expected size of the data (if known) • Outline the data utility: to whom will it be useful
2. FAIR Data 2.1. Making data findable, including provisions for metadata	• Outline the discoverability of data (metadata provision) • Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? • Outline naming conventions used • Outline the approach towards search keyword • Outline the approach for clear versioning • Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
2.2 Making data openly accessible	• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so • Specify how the data will be made available • Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? • Specify where the data and associated metadata, documentation and code are deposited • Specify how access will be provided in case there are any restrictions
2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. • Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary

	interoperability? If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	• Specify how the data will be licenced to permit the widest reuse possible • Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed • Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why • Describe data quality assurance processes • Specify the length of time for which the data will remain re-usable
3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your project • Describe costs and potential value of long term preservation
4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)