

30-06-2024

Deliverable D4.2: Data Management Plan

Deliverable D4.2

Contractual Date: 30-06-2024
Actual Date: 30-06-2024
Grant Agreement No.: 101123289
Work Package: WP4
Task Item: T4.2
Lead Partner: EITD

Authors: Andrea Biancini (EITD)

Abstract

This Data Management Plan (DMP) outlines how data within the project is managed in a FAIR (findable, accessible, interoperable, and reusable) manner. The DMP has been created in accordance with the Horizon Europe Data Management Plan Template. The DMP should include information on the handling of research data during and after the end of the project, what data will be collected, processed, and/or generated, how data will be curated and preserved, and other relevant issues. By following the guidelines outlined in the DMP, project partners can ensure that data is managed in a way that maximizes its value and impact.

© EIT Digital on behalf of the EMAI4EU project.

The activities leading to these results has received funding from the European Community's DIGITAL Programme under Grant Agreement No. 101123289 (EMAI4EU).

This document is licensed under a [Creative Commons Attribution 4.0 license](https://creativecommons.org/licenses/by/4.0/)



Versioning and contribution history

Version	Date	Authors	Notes
0.1	29/05/2024	Andrea Biancini (EITD)	Draft version.
0.2	10/06/2024	Javier Segovia & Mayte Sanchez (UPM)	Minor formatting changes.

Table of Contents

1	Introduction.....	5
1.1	EMAI4EU.....	5
1.2	Work Package 4.....	6
1.3	Deliverable 4.2.....	7
1.3.1	Purpose.....	7
1.3.2	Objectives.....	7
2	Data Summary.....	7
3	FAIR data.....	10
3.1	Making data findable, including provisions for metadata.....	10
3.2	Making data openly accessible.....	11
3.3	Making data interoperable.....	12
3.4	Increase data re-use (through clarifying licences).....	12
4	Allocation of resources.....	13
5	Data security.....	14
6	Ethical aspects.....	14
7	Other issues.....	15
	Appendix A – FAIR Data Management at a glance.....	Errore. Il segnalibro non è definito.

Table of Figures

No table of figures entries found.

Table of Tables

Table 1: Data sets overview.....	8
Table 2: Data sets description and utility	10
Table 3: Data sets accessibility	12
Table 4: Issues covered in the EMAI4EU FAIR Data Management Plan	Errore. Il segnalibro non è definito.

1 Introduction

This Data Management Plan has been created following the Horizon 2020 FAIR DMP template, which has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data.

This document is intended to follow the best practices for a FAIR data management¹.

Definition: FAIR data management

In general terms, your research data should be 'FAIR', that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution.

This Data Management Plan is a set of questions, from the Horizon 2020 template, that were answered with a level of detail appropriate to the project. This DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur.

As a minimum, the DMP should be updated in the context of the periodic evaluation/assessment of the project. If there are no other periodic reviews envisaged within the grant agreement, an update needs to be made in time for the final review at the latest.

1.1 EMAI4EU

EMotion Artificial Intelligence specialists for Europe (EMAI4EU) aims to train the next generation of specialists and innovators in Emotion Artificial Intelligence in Europe. EMAI4EU will achieve this goal by designing and delivering a double-degree master's programme (ISCED Level 7, 120 ECTS) in Artificial Intelligence with a specialisation in Emotion Artificial Intelligence and a minor in Innovation and Entrepreneurship. The master's programme will be designed and delivered by 8

¹ FAIR Data Principles (FORCE11 discussion forum): <https://force11.org/group/fairgroup/fairprinciples>

FAIR principles (article in Nature): <https://www.nature.com/articles/sdata201618>

higher education institutions from 5 different countries with 4 innovative SMEs, a leading research centre in AI and EIT Digital, a pan-European organisation with experience in delivering education programmes in advanced digital skills across Europe. EMAI4EU master's programme will foster strong interactions and mobility between academia and business, strengthen knowledge triangle integration, promote entrepreneurship, foster inclusiveness, and boost the growth of the existing EIT Digital ecosystem, one of the largest digital ecosystems in Europe. In addition to the specialised master's programme, EMAI4EU partners will develop and deploy self-standing learning modules on topics related to Artificial Intelligence and Emotion Artificial Intelligence. These modules will result in a set of certifications in advanced digital skills released by participating higher education institutions and EIT Digital. In line with the goals of the Digital Compass and New European Innovation Agenda, EMAI4EU will train more than 1000 participants across four years and contribute to reducing the gap in advanced digital skills in Europe and increase Europe's competitiveness in a key digital technology domain such as Artificial Intelligence. Finally, it is important to highlight the added value of this proposal: Several isolated courses on Emotion AI can be identified in some European universities, but a master's programme in AI with a specialisation in Emotion AI is not available on the EU market.

1.2 Work Package 4

The objectives of Work Package 4 are:

- To ensure the overall management of the project and effectively monitor the project, in administrative, technical, and financial terms.
- To guarantee high-quality content and management with the aim of securing effective progress.
- To coordinate the enrolment process of participants to EMAI4EU education programmes.
- To ensure the establishment of effective and sustainable partnerships within the consortium.

It is concerned with undertaking the technical and scientific coordination of the EMAI4EU project as well as the administrative and financial management. This work package will ensure that appropriate quality control and reporting mechanism are applied across the project.

1.3 Deliverable 4.2

1.3.1 Purpose

The EMAI4EU Data Management Plan has been prepared with these two purposes:

1. to describe the data management life cycle for the data to be collected, processed and/or generated by the EMAI4EU project;
2. include information on the handling of research data during and after the end of the project, what data will be collected, processed and/or generated, how data will be curated and preserved, and resource and budgetary planning for data management.

1.3.2 Objectives

- Ensure effective management of research data throughout the project life cycle.
- Describe the data management life cycle for the data to be collected, processed and/or generated by the project.
- Ensure that research data is findable, accessible, interoperable and re-usable (FAIR).
- Ensure that research data is managed in compliance with the General Data Protection Regulation (GDPR).
- Reflect the current state of consortium agreements on data management and be consistent with exploitation and Intellectual Property Rights (IPR) requirements.
- Provide an overview of all datasets collected and generated by the project and define the consortium's data management policy and approach.

2 Data Summary

In order to provide an overview of the different data sets that are currently and will be produced in the EMAI4EU project, we need to distinguish two types of data:

1. Non-sensitive data produced by the project and released for potential reuse in other projects or research activities.
2. Operational data used to implement the activities described in the project. This data includes very frequently also sensitive data about students and participants to training activities.

The following table shows the data type, the origin of the data, the related WP number and the format, in which the data will be presumably stored.

#	Data type	Type	Origin	WP#	Format
1	Market review of Artificial Intelligence sector.	Non-sensitive	Derived data by other reports and market data.	WP1	.docx
2	Literature review data on Artificial Intelligence..	Non-sensitive	Derived data by publications or published reports.	WP1	.docx
3	Recruitment cycle data about participants.	Operational	Primary data	WP1, WP2	.xlsx and .docx
4	Personal data of students participating to master programs.	Operational	Primary data	WP1, WP2	.xlsx and university systems
5	Data on participants to self-standing modules.	Operational	Primary data	WP1, WP2	.xlsx
6	Satisfaction survey from students at the end of a learning course or activity.	Operational	Primary data	WP1, WP2	.xlsx
7	Marketing data related to communication and dissemination activities.	Operational	Primary data	WP3	.xlsx

Table 1: Data sets overview

Table 2 describes the data set and the purpose of the data collection of data generation in relation with the objectives of the project. Additionally, it shows the data utility for clarifying to whom the data might be useful.

#	Data type	Description & Purpose	Utility
1	Market review of Artificial Intelligence sector.	<p>Description The data contains the result of a market review analysis done on the field of artificial intelligence. The analysis will be performed by analyzing publicly available market data and by interviewing economic actors in the sector.</p> <p>Purpose The collection of this data will serve as an input to the process of review of the master program curriculum. This data will also serve to guide the definition of the content for the self-standing learning modules.</p>	<p>The data could be useful for research on the artificial intelligence sector.</p> <p>It can also be useful for other educational institutions and to organizations and business to better understand the current state of the market, identify the latest trends and threats and make informed decisions about artificial intelligence products and services.</p>

#	Data type	Description & Purpose	Utility
2	Literature review data on Artificial Intelligence.	<p>Description The data contains the result of a literature review done on the field of artificial intelligence. The analysis will be performed by analyzing publications, articles and course syllabus from other universities and higher education institutions.</p> <p>Purpose The collection of this data will serve as an input to the process of review of the master program curriculum on Emotional AI. This data will also serve to guide the definition of the content for the self-standing learning modules.</p>	<p>The data could be helpful to researcher interested in understanding the current state of knowledge in the field of artificial intelligence, identify gaps in the literature and develop research question and hypotheses.</p> <p>The data can also be useful for policymakers by helping the development of policies and regulations that are evidence-based and effective.</p>
3	Recruitment cycle data about participants.	<p>Description This data includes all the personal information of candidates applying for the master programmes. The data will include contact information, CV history and study track records for all applicants.</p> <p>Purpose Data is gathered for administrative purposes and to enable the selection of candidates, including the awarding of scholarships, based on their recent educational and professional history.</p>	<p>The data could be helpful specifically to employers and recruitment agencies to make informed decisions about which candidates to hire and which positions to offer.</p> <p>Researchers can use this data to study the qualifications and backgrounds of candidates applying for master's programs or jobs. The data can help researchers identify trends and patterns in the qualifications and backgrounds of successful candidates and develop research questions and hypotheses.</p>
4	Personal data of students participating to master programs.	<p>Description This data includes all the personal information of students of the master programs. The data will include contact information, CV history and study track records for all students and will be managed by the guesting universities following the general rules for all students.</p> <p>Purpose Data is gathered for administrative purposes and to enable participation to the courses and track of the student's path.</p>	<p>The data could be helpful specifically to employers and recruitment agencies to make informed decisions about which candidates to hire and which positions to offer.</p> <p>Researchers can use this data to study the qualifications and backgrounds of learners of the self-standing modules. The data can help researchers identify trends and patterns in the qualifications and backgrounds of online students and develop research questions and hypotheses.</p>
5	Data on participants to self-standing modules.	<p>Description Data related to the registration and participation to self-standing modules. This data includes contact information and digital addresses of all participants. The</p>	<p>The data could be helpful specifically to employers and recruitment agencies to make informed decisions about which</p>

#	Data type	Description & Purpose	Utility
		<p>data also includes information regarding eventual certifications obtained by the participants.</p> <p>Purpose Data is gathered for administrative purposes to enable the access to the online platform and the tracking of the study activities.</p>	<p>candidates to hire and which positions to offer.</p> <p>Researchers can use this data to study the qualifications and backgrounds of learners of the self-standing modules. The data can help researchers identify trends and patterns in the qualifications and backgrounds of online students and develop research questions and hypotheses.</p>
6	Satisfaction survey from students at the end of a learning course or activity.	<p>Description Data related to the results of the satisfaction survey gathered from students of master and self-standing modules.</p> <p>Purpose Data is gathered to implement a quality improvement process and to improve courses and training material.</p>	The data is of interest to the project participants to obtain workable feedbacks and aliment a continuous improvement cycle to improve the course material and training paths.
7	Marketing data related to communication and dissemination activities.	<p>Description Data regarding the communication and dissemination campaign on social networks and digital channels.</p> <p>Purpose Digital marketing will be a central part of the strategy of attraction to candidate students to the project's program. Collecting operational data is fundamental for digital communication to work effectively.</p>	<p>The data could be helpful to digital marketing agencies or marketing professional interested in evaluating the effectiveness of their digital marketing campaigns and identify areas for improvement.</p> <p>Researchers can use this data to study the effectiveness of digital marketing campaigns and identify trends and patterns in the participation of potential students or customers.</p>

Table 2: Data sets description and utility

3 FAIR data

3.1 Making data findable, including provisions for metadata

To ensure that the data generated during the project is findable, we will implement the following provisions:

- All data will be recorded in a predeterminate structure and with an agree format.
- Data structure and format will ensure interoperability and ease of use.

To ensure that the data is discoverable, we will implement the following mechanisms:

- Data will be assigned unique identifiers to enable easy identification and tracking.
- Data will be stored in a structured and organized manner to enable efficient searching and browsing.
- Data will be made available through appropriate repositories and archives to enable discovery and reuse.

3.2 Making data openly accessible

The following table is highlighting which data that are produced and used in the project and which will be made openly available. It also explains why several datasets cannot be shared because of particular reasons and, in this case, an alternate solution will be presented.

#	Data type	Openly available	Justification	Alternate solution
1	Market review of Artificial Intelligence sector.	Yes	Results of this analysis will be described in the project deliverable D1.1.	<i>(not relevant)</i>
2	Literature review data on Artificial Intelligence.	Yes	Results of this review will be described in the project deliverable D1.1.	<i>(not relevant)</i>
3	Recruitment cycle data about participants.	No	The sensible data about students involved in the recruitment process of the project will not be released, in respect to GDPR and any other regulation that may apply.	Statistical data about the recruitment cycle and student admissions will be described in the project deliverables D4.4, D4.5, D4.6, D4.7.
4	Data of students participating to master programs.	No	The sensible data about students participating to courses and learning activities of the project will not be released, in respect to GDPR and any other regulation that may apply.	Statistical data about the student participation to the master will be described in the project deliverables D1.2, D1.3, D1.4.
5	Data on participants to self-standing modules.	No	All personal data, including contact information, regarding students, will not be make openly available, in respect to GDPR and any other regulation that may apply.	Statistical data about the student participation to the master will be described in the project deliverables D2.2, D2.3.

#	Data type	Openly available	Justification	Alternate solution
6	Satisfaction survey from students at the end of a learning course or activity.	No	All personal data of students, including contact information and opinion on the course quality, will not be made openly available, in respect to GDPR and any other regulation that may apply.	Statistical data about the student satisfaction expressed for the courses attended will be described in the project deliverables D1.2, D1.3, D1.4, D2.2, D2.3.
7	Marketing data related to communication and dissemination activities.	No	The granular and analytical marketing data used to guide communication and dissemination activities will not be released.	Statistical aggregated data about marketing and dissemination activities will be released in deliverables D3.2, D3.3, D3.4, D3.5.

Table 3: Data sets accessibility

The data made available, will be registered in official project deliverables and, as such, will be published on the project website and on the EC portal for public access.

3.3 Making data interoperable

All the data shared by the project will use document standards that will make it interoperable. The data will be released with docx or xlsx file formats promoting interoperability by providing a standardized, open, and flexible way to exchange and reuse data across different systems and applications.

3.4 Increase data re-use (through clarifying licences)

To permit the wider re-use of data, all openly available project deliverables and main results will be released with a Creative Commons Attribution (CC-BY) license. This license allows others to distribute, remix, and build upon the data, even commercially, as long as they credit the original source.

The data released under this license does not include:

- any sensible information regarding students, that will be protected by GDPR and other relevant regulation,
- the master course, that will remain property of the producing entity, and

- the online training modules, that will maintain a shared ownership between the beneficiaries that have generated them.

4 Allocation of resources

The following resources will be allocated to ensure effective data management throughout the project:

1. **Personnel:** Data management will be overseen in Task 4.5 of the WP4. The task will permit all partners involved in these activities to dedicate resources, including personnel, to the tasks and activities related to data management. A data manager will be appointed to oversee the implementation of the data management plan and ensure compliance with relevant regulations and guidelines. The data manager will be responsible for creating and maintaining the metadata, ensuring data quality, and managing the storage and security of the data. The data manager will also be responsible for training project personnel in data management best practices.
2. **Infrastructure:** The project will allocate resources for the storage and backup of data in secure locations. For this purpose, the project will use the website and the Teams instance of the coordinator partner: EIT Digital.
3. **Budget:** The project will allocate a budget for data management activities. The budget will also include provisions for the dissemination and sharing of data, including the use of appropriate repositories and archives.
4. **Deliverables:** This deliverable D4.2 of the project describes an initial data management plan. The project will also include deliverables for the dissemination and sharing of data.

The allocation of resources will be reviewed and updated throughout the project as necessary to ensure that the data management plan remains effective and compliant with relevant regulations and guidelines. The project will also ensure that the allocation of resources is consistent with exploitation and Intellectual Property Rights (IPR) requirements.

5 Data security

The project will implement the following measures to ensure the security of the data recorded on the Microsoft Teams platform used to store all relevant project data:

1. **Access controls:** Access to the data will be restricted to authorized personnel only. The project will use Microsoft Teams to manage access controls, including role-based access controls and multi-factor authentication.
2. **Backup and recovery:** The project will implement a backup and recovery plan to ensure that the data is recoverable in the event of a disaster or system failure. The project will use the Microsoft's platform backup and recovery capabilities to ensure that the data is protected.
3. **Data retention and disposal:** The project will implement a data retention and disposal policy to ensure that data is retained only for as long as necessary and disposed of securely when it is no longer needed.
4. **Monitoring and auditing:** The project will implement monitoring and auditing procedures to ensure that the data is being used appropriately and that any unauthorized access or use is detected and addressed.

The project will also ensure that all personnel involved in the project are trained in data security best practices and that they understand their roles and responsibilities in protecting the data. The project will also ensure that all data management activities are compliant with relevant regulations and guidelines.

6 Ethical aspects

The project will ensure that all data management activities are conducted in compliance with relevant ethical guidelines and regulations. The following ethical aspects will be considered:

1. **Informed consent:** The project will obtain informed consent from all participants before collecting any data. Participants will be informed about the purpose of the data collection,

how the data will be used, and any potential risks or benefits associated with the data collection.

2. **Data privacy:** The project will ensure that all data is collected, stored, and shared in compliance with relevant data privacy regulations. The project will implement appropriate measures to protect the privacy and confidentiality of the data, including encryption, access controls, and data anonymization where necessary.
3. **Data ownership:** The project will ensure that all data is owned by the appropriate parties and that any intellectual property rights are respected. The project will also ensure that any data sharing or dissemination is conducted in compliance with relevant regulations and guidelines.
4. **Data sharing:** The project will ensure that any data sharing or dissemination is conducted in compliance with relevant regulations and guidelines. The project will also ensure that any data sharing or dissemination is conducted in a manner that respects the privacy and confidentiality of the data.
5. **Data retention and disposal:** The project will implement a data retention and disposal policy to ensure that data is retained only for as long as necessary and disposed of securely when it is no longer needed.

7 Other issues

Data management in the project will be performed following the European Commission's Horizon 2020 procedures. In particular this document represents the Data Management Plan (DMP) as requested in the program and describes the data management life cycle for the data to be collected, processed, and/or generated by the project.

References

- [DIGITAL]** <https://digital-strategy.ec.europa.eu/en/activities/digital-programme>
- [EMAI4EU]** <http://eitdigital.eu/emai4eu/>

Glossary

Community	A group of users, organised with a common purpose, and jointly granted access to resources. It may act as the interface between individual users and the resources. (see also [WISE-SCI])
AI	Artificial Intelligence
EC	European Commission
EIT	European Institute of Innovation and Technology
KIC	Knowledge and Innovation Community
GA	Grant Agreement
GDPR	General Data Protection Regulation
R&S	Research and scholarship