



## **D2.3 – Initial data management plan**



<b>Project Title</b>	A network of excellence for distributed, trustworthy, efficient and scalable AI at the Edge
<b>Project Acronym</b>	dAIEDGE
<b>Grant Agreement No</b>	101120726
<b>Topic</b>	HORIZON-CL4-2022-HUMAN-02-02
<b>Start Date of Project</b>	September 1 <sup>st</sup> , 2023
<b>Duration of Project</b>	36 Months

<b>Name of the Deliverable</b>	Initial data management plan
<b>Number of the Deliverable</b>	D2.3
<b>Related WP Number and Name</b>	WP2 - Project Management and Coordination
<b>Related Task Number and Name</b>	T2.3 - Data management, quality control and risk monitoring
<b>Deliverable Dissemination Level</b>	PU – Public
<b>Deliverable Due Date</b>	29.02.2024
<b>Deliverable Submission Date</b>	29.02.2024
<b>Task Leader/Main Author</b>	DFKI
<b>Contributing Partners</b>	
<b>Reviewer(s)</b>	DFKI

**Keywords**

Data management plan

## Revisions

Version	Submission date	Comments	Author
v0.1	22.02.2024	Initial version	DFKI
v0.2	28.02.2024	Internal review	DFKI
v1.0	29.02.2024	Coordinator review	DFKI

## Disclaimer

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

## Acronyms and definitions

Acronym	Meaning
FSTP	Financial Support to Third Parties
FAIR	Findable, Accessible, Interoperable, Reuseable

## ***Executive Summary***

This deliverable presents the data management plan of the project dAIEDGE. The type of data processed in the project is described, along with the measures to ensure that the data is FAIR (findable, accessible, interoperable and reusable). The open sciences practice and the procedures for data security and GDPR compliance are detailed. Our concept for the storage and sharing of dAIEDGE code in dedicated repositories is presented.

## CONTENTS

1. Introduction .....	6
2. Data summary.....	6
3. Dissemination of research results .....	7
4. Open Science practices .....	7
5. FAIR data.....	8
5.1. Making data findable .....	9
5.2. Making data openly accessible .....	9
5.3. Making data interoperable .....	10
5.4. Increase data re-use.....	10
6. dAIEDGE code repository.....	10
7. IP Protection .....	11
8. Data security and GDPR compliance .....	11

## 1. Introduction

This deliverable describes the approach followed for project-related data management and the functionalities of the common project data platforms.

## 2. Data summary

This project uses data at various levels, which will be gathered by project partners. The following types of data are foreseen to be collected in the project:

- Personal text data (Person names, e-mail addresses, companies etc): for stakeholder mapping to find appropriate participants for workshops, through surveys and workshops to evaluate the outreach activities, potential FSTP applicants and materials produced.
- Users feedback for evaluation purposes: interview protocols conducted with specific stakeholders and users, qualitative and quantitative feedback of participants.
- Metadata on methodologies, instruments, procedures, the research goal and its target groups will be collected.
- Data collected for learning-based algorithms from public datasets or newly created dataset.
- Data stored by the dAIEDGE Virtual Lab framework or the dAIEDGE Marketplace engine (training data, test data, algorithms, parameters, metadata).
- Research result in terms of
  - Software
  - Scientific finding (e.g., paper)
  - Evaluation results

The collected data might be useful to other researchers who are working in related fields, edge AI and broadly AI, scientific communication, education and other research fields. This will be achieved by providing open access to the research data according to the description in section 4.

### 3. Dissemination of research results

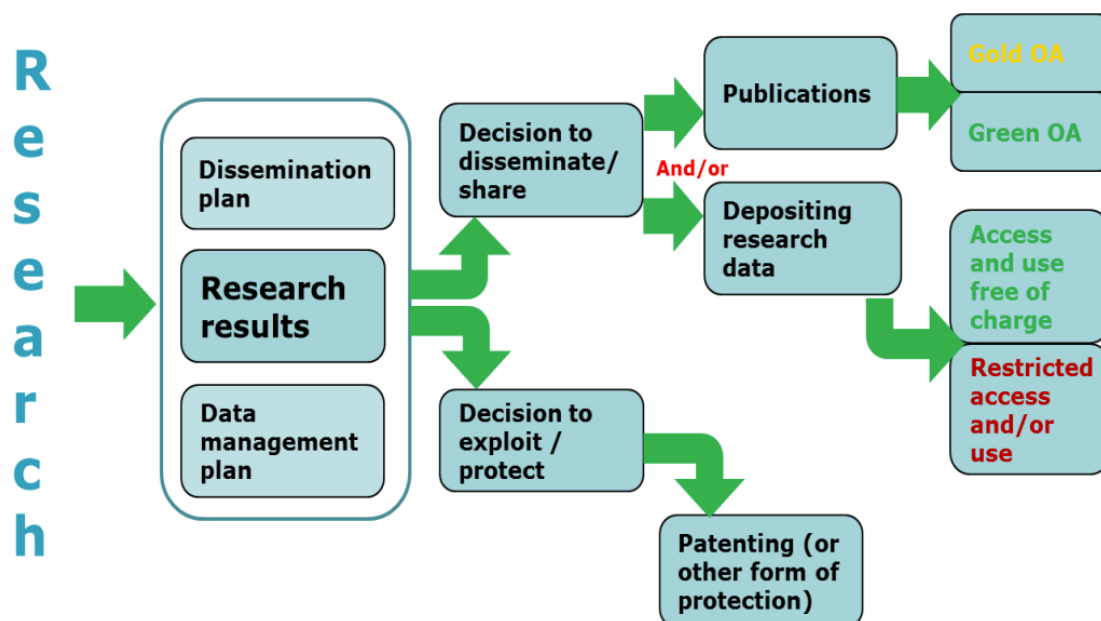


Figure 1: Open access decision graph (from "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020")

For all data types, the Consortium will analyse if there are any (potential) conflicts against commercialization, privacy or IPR protection issues on knowledge being generated. The decision process in Figure 1 summarizes how the consortium decides to make information publicly available and when.

### 4. Open Science practices

The dAIEDGE consortium will participate in the Open Research Data Horizon Europe policy. The Open Science (OS) strategy considered will align with EC objectives for Responsible Research and Innovation, defining reliable project management procedures and bodies (Data Management Plan, Steering Committee) to guarantee proper OS practices application for all research outputs (such as publications, data, software, models, algorithms, and workflows) is considered throughout the project duration. dAIEDGE will follow several practices that have proven to bear merit for improving research quality and impact:

1. Publish as open as possible: For instance, IEEE, ACM and USENIX venues and 'Open Research Europe' platform. In addition, as many deliverables as possible from the project will be publicly available.
2. Early release: dAIEDGE will create a dedicated channel in a public repository (e.g., GitHub, or the dAIEDGE Virtual Lab) for its open source/APIs releases to optimise exchange of information, including intermediate releases for early user access and feedback. Access to the different modules will be available from the project webpage. It will be a tool used for the FSTP, to exchange information with the selected participants.



3. Research output management: In general, a combined scheme is expected, where results are released as open as possible and as close as necessary.
4. Reproducibility of results: All research outputs will include detail on how evaluation is performed. Some venues have introduced a results reproducibility certificate. When applicable, dAIEDGE results will use this process.
5. Open peer review: In order to enhance transparency and encourage interaction, manuscripts will be made available (e.g. arXiv, Open Research Europe) in advance of any formal peer review procedures. dAIEDGE will ensure protection of knowledge by adopting licenses which enable free circulation of documents while safeguarding author and project IPR. In addition, dAIEDGE will pay attention to protecting the project and EC image (e.g., against de-contextualization) by ensuring content integrity.

Furthermore, the following specific open science practices are implemented as integral parts of our overall methodology and approach:

- Sharing proposals for open standards, especially for the inclusion of third-party modules using open API to our platform
- Sharing Open Educational Resources, such as demonstrators, documents, videos, code repositories
- Sharing Training and Testing datasets, in particular for machine learning algorithms and AI algorithms that will be used in the dAIEDGE framework and published in conferences and journals.

dAIEDGE academic partners will also share their research findings in early stages through preprint platforms such as arXiv. The formal presentation of scientific results will be made available to the research community through open access journals. All open-access publications will be also deposited on the open repository ZENODO<sup>1</sup>, developed by OpenAIRE, on which a dAIEDGE community has been created. These publications will be linked to corresponding original research data (and metadata), further simplifying access to and/or use of dAIEDGE data for external researchers and other third parties.

## 5. FAIR data

dAIEDGE follows the open research data concept, this implies that all research data should be “FAIR”, that is findable, accessible, interoperable and re-usable.

In the project all data will follow a clear version number structure, if needed. For all quantitative and qualitative research in the project, non-identifiable metadata will be produced and made available on the dAIEDGE internal SharePoint platform being accessible by all members of the consortium. Metadata will describe instruments used, methodologies employed and goals and

---

<sup>1</sup> <https://zenodo.org/communities/daiedge>

target groups of the research. Metadata will be collected and appropriately stored by the researchers. All data will be anonymized prior to uploading to the repositories.

All data produced and the associated anonymised metadata, documentation and code will be stored on the internal dAIEDGE repository. For making results publicly available, we will use Zenodo and provide an overview of the available data on the project website. For data that will be shared with a selected audience (e.g., the project consortium), we recommend using the detailed access control of Zenodo. It enables interested researchers to request access to restricted data directly from the owner of the data and it allows the owner of the data to define and revoke access rights per request. To allow data exchange and re-use between researcher, institutions, organisations, countries, etc. the dAIEDGE project will assure the use of interoperable formats. Moreover, standard vocabularies for all data types will be used to allow inter-disciplinary interoperability. In the case where less common ontologies or vocabularies cannot be avoided or are specific to the project itself, dAIEDGE will provide mappings to more commonly used ontologies.

The following paragraphs gives some more Details on how the FAIR principles will be implemented in dAIEDGE.

### **5.1. Making data findable**

Publicly shared data produced within the project must be findable. Zenodo provides the needed search functionality that allows filtering by keywords and the community's name. It's up to the partner to provide appropriate keywords during the data publication process. These keywords should assist people that search for the data at Zenodo. Each upload receives a digital object identifier (DOI) which will make it easily findable even though the URL to the dataset might have changed. The bibliography metadata should additionally include the following:

- The terms “European Union (EU)” and “Horizon Europe”
- The name of the action, acronym and grant number
- The publication date, length of embargo period

All partners are encouraged to also upload their restricted data to Zenodo. Zenodo allows restricted access for confidential data. Note that the data is still findable although it is not visible if the access is restricted.

All internal documents, like minutes, media, confidential deliverables will be archived by DFKI on their SharePoint repository being accessible by all project partners. Documents will be findable through the hierarchical directory structure (e.g., ordered by categories like deliverables, datasets, meetings, media, press).

### **5.2. Making data openly accessible**

Not all produced data can be made openly accessible. The consortium will discuss in each case if data generated by a partner can be made available to the public or should be kept confidential for internal use.

Nevertheless, important scientific data that is needed to reproduce or validate results must be made openly accessible and any publications must be published open access.

All data should be accessible with standard software, with a preference for free standards (.pdf, .tex, .jpg, .png, ASCII, etc.). Proprietary data formats should be avoided at any time. According to the use case, or type of data converting to standard formats should be considered over publishing proprietary data.

### 5.3. Making data interoperable

Controlled vocabulary should be used according to the target audience of a public document. If the published data is a dataset or source code, a proper documentation is expected including an example application / script to make it easy to reuse the data. If free software was used for generating the result it should also be considered to attach a Dockerfile to the dataset or source code to avoid dependency issues due to constantly evolving operating systems and third-party libraries changing their API.

### 5.4. Increase data re-use

All data and documents being published need to be published with a license. The partner itself must make several decisions that affect the choice of an appropriate license:

- Is commercial use permitted?
- Is modification permitted?
- Is distribution permitted?
- Is private use granted?
- Is redistribution mandatory?

The licenses that fit best can be selected on the Creative Commons webpage. Other types of licenses are available for software. The choice of license depends on the specific requirements of the copyright owner. In case software should be released as open-source partners can generally choose between licenses that allow commercial use (e.g., BSD) and those that do not (e.g., GNU General Public License, GPL).

Public data will be made available as early as possible. The data should be produced, cleaned and documented prior to making them available, if there's no embargo period on the corresponding publication and its need to be published when the publication has been published, at latest at the end of the project.

## 6. dAIEDGE code repository

The software developed for the dAIEDGE project by the partners will be made accessible over two mechanisms developed in the project lifetime: the dAIEDGE Virtual Lab platform and the dAIEDGE Marketplace. The results of dAIEDGE will be also be made available on the EU AlonDemand platform whenever possible.

**The dAIEDGE Virtual Lab** serves as a unifying platform, connecting existing and forthcoming research infrastructures through a federated architecture and a suite of interoperable connectors. Its primary objective is to facilitate remote access to a wide array of AI resources, ranging from laboratory equipment for conducting experiments to vast datasets essential for training AI algorithms. The dAIEDGE Virtual Lab will be integrated with established platforms such as the European AI-on-demand (AIOD) platform, the Bonseyes AI Marketplace, and the ARIAC/CyberExcellence factory. By leveraging these platforms, the Virtual Lab will offer remote access to valuable resources necessary for AI at the edge. The Virtual Lab is envisioned as a collaborative endeavor, initially comprising contributions from the dAIEDGE consortium partners. The assets will not be limited to digital resources, but hardware components such as RPI4, Jetson Nano, NX, STM MP1/H7, GAP8, and others provided by the dAIEDGE project will also be made available for experimentation. In terms of data protection and accessibility, the resource management and more generally access rights within the Virtual Lab will be governed by licenses and will be based on blockchain technology to ensure transparency and accountability.

**The dAIEDGE Martekplace** will serve as a bridge between research and business development, as it will facilitate the creation and exchange of AI applications and services optimized for low-power processors, thereby diminishing the carbon footprint associated with AI computing. Additionally, it will expand access to a distributed community of AI talents, empowering them to tackle enterprise challenges. It will also enable low-tech SMEs to access and utilize AI services, while concurrently addressing governance concerns through a mutualized governance and user incentive model driven by token ownership, aligning with European objectives for trustworthy AI. It will be develop as a Blockchain-based and distributed version of the distributed Bonseyes AI Marketplace, ensuring that data exchange is secure and traceable.

**Project GitLab** – In the case where software developed by dAIEDGE partners have to be shared internally, this code can be stored on a GitLab platform managed by the coordinator DFKI. A GitLab group “dAIEDGE” has been created, where partners can create repositories for GitLab projects. This GitLab platform has an integrated Continuous Integration mechanism to ensure the quality of the code pushed to master branches of the projects.

## 7. IP Protection

In the dAIEDGE Consortium Agreement (CA), the background-IP that each partner brings into the project is clearly stated in order to ensure its protection. The consortium agreement also defines specific rules according to which any disputes regarding IP that might arise will be resolved.

## 8. Data security and GDPR compliance

Every project partner organization is responsible for ensuring data security for all dAIEDGE related data.

General Data Protection Regulation (GDPR) compliance is enforced at several levels. Initially, partners are responsible for following the GDPR whenever data is acquired. The persons involved in

such experiments in the project, are responsible and agree to seek guidance and obtain permission from the data protection officers of their organizations whenever person-related data is collected or made public. In addition to that, there will be general supervision of all data collection activities by the internal project ethics officer as well as the external ethics advisor of dAIEDGE.



**[daiedge.eu](https://daiedge.eu)**