



CONVOLVE

Seamless design of smart edge processors

GRANT AGREEMENT NUMBER: 101070374

Deliverable D9.2

Data Management Plan

Title of the deliverable	Data Management Plan
WP contributing to the deliverable	WP 9
Task contributing to the deliverable	Task 9.3
Dissemination level	PU – Public
Due submission date	30/04/2023
Actual submission date	30/04/2023
Author(s)	Sander Stuijk (TUE), Victor Sanchez (TUE)
Internal reviewers	Tobias Piechowiak (GNA)

Document Version	Date	Change
V0.1	17/03/2023	Initial ToC
V1.0	30/04/2023	First complete version DMP

Table of Contents

Table of Contents	3
Deliverable Summary	4
1. Introduction	4
1. Data Summary	5
2. FAIR data	5
2.1. Making data findable, including provisions for metadata	5
2.2. Making data accessible	6
2.3. Making data interoperable	6
2.4. Increase data re-use	6
3. Other research outputs	7
4. Allocation of resources	7
5. Data security	7
6. Ethics	8
7. Other issues	8

Deliverable Summary

This document presents the Data Management Plan of the CONVOLVE project. It identifies the datasets to be produced in this project as well as the principles used to ensure the data is curated according to the FAIR principles (Findable, Accessible, Interoperable and Reusable).

1. Introduction

The scope of the first version of the CONVOLVE Data Management Plan (DMP) for the EU is to define key elements that will facilitate the potential reuse of the data collected and processed within the project. The DMP will ensure that this data will be findable, accessible, interoperable and reusable (FAIR), in accordance with the requirements for research data management of Horizon Europe as described in article 17 (annex 5) of the grant agreement.

1.1 Project background

The CONVOLVE project is funded by the European Commission under its Horizon Europe programme. CONVOLVE enforces the EU's position in the design and development of smart edge-processors, such that it can become a dominant player in the global edge processing market.

With the rise of deep learning (DL), our world braces for Artificial Intelligence (AI) in every edge device, creating an urgent need for edge-AI processing hardware.

Unlike existing solutions, this hardware needs to support high throughput, reliable, and secure AI processing at ultra-low power (ULP), with a short time to market.

With its tradition in edge solutions and open processing platforms, the EU is ideally positioned to become the leader in this edge-AI market. However, certain demands need to be met to position the EU as a leader in these technologies: Edge processors need to become 100x more energy efficient; Their complexity demands automated design with 10x design time reduction; They must be secure and reliable to get accepted; Finally, they should be flexible and powerful to support the DL domain. CONVOLVE addresses these demands and thereby enables EU leadership in Edge-AI.

1.2 Data Management Plan objectives

This document is the initial CONVOLVE Data Management Plan (DMP). The DMP identifies the datasets that will be generated and used during the project and defines each dataset's purpose and life cycle.

In accordance with Guidelines on FAIR Data management, the objectives of the DMP are:

- To identify the datasets that CONVOLVE will produce;
- To define how these datasets will be made 'FAIR' (Findable, Accessible, Interoperable and Reusable);
- To define the allocation of resources (costs and responsibility) for data management during and after the project;

- To define procedures for data security (including data recovery as well as safe storage) during the project and for long term preservation.

The Deliverable 9.2 “Data management plan” is part of work package 9 “Coordination and Management”. The DMP is intended to be a living document where information will be continuously added and revised as the implementation progresses.

As a key element of sustainable data management, the DMP outlines what, and how, data will be collected, processed and/or generated, which methodology and standards will be applied, whether data will be shared/made open access and how data will be curated and preserved during and after the end of the project. It provides guidelines, answers to issues, and presents the approach that the CONVOLVE project will adopt with respect to the management and protection of data. It is the responsibility of each partner to notify the coordinator of changes in the data they are collecting during the project.

2. Data Summary

The CONVOLVE project will use existing public domain datasets (e.g., ImageNet, CIFAR100) to train neural networks developed in this project. It is expected that the project will not generate any new datasets.

3. FAIR data

This section is based on the standard Horizon Europe template for FAIR data. The beneficiaries have agreed on the use of the Zenodo platform to share any data marked as public domain. A special CONVOLVE community (zenodo.org/communities/convolve) has been created on this platform to make public CONVOLVE output easily findable and accessible.

3.1 Making data findable, including provisions for metadata

The Zenodo platform is utilised for public-domain validation of research findings. Along with the data, metadata will be produced before uploading on Zenodo, and the DOI assigned by Zenodo will allow data to become findable.

Data will be described by rich metadata using standard or domain specified terminologies. It is not expected that new terminology will be developed as part of this project.

TUE, as coordinator, will maintain the Convolve community on the Zenodo platform, uploading data for the consortium. Other partners can be provided with upload access to the community pages on request.

Public deliverables and promotional materials (such as leaflet of events) will be made available through the CONVOLVE webpage (convolve.eu). Whereas confidential research data will not be made findable, and only upon consortium partners’ request will be disclosed with metadata.

3.2 Making data accessible

3.2.1 Repository

The Zenodo platform will be used to deposit any output marked as public domain. A special CONVOLVE community (zenodo.org/communities/convolve) has been created on this platform to make public CONVOLVE outputs easily findable and accessible. Moreover, Zenodo will assign a DOI to all datasets uploaded to the platform which enables reliable identification and a uniform access method to the datasets.

3.2.2 Data

Data publicly accessible are:

- Non-confidential data needed to validate the results presented in scientific publications and non-confidential products of research created in the project (on the Zenodo platform).
- Public deliverables and promotional materials created in CONVOLVE are made available for download from the project website (convolve.eu).

All the other data generated, collected, used, or processed during the project development will be considered confidential as default and a part of this will be stored according to the processes described in the project handbook (Deliverable D9.1).

The Zenodo policy for deposition is transparent and available on its website and no additional arrangements have been made so far. If required, an “embargo strategy” will be implemented to minimize restrictions on sharing. The process to seek protection of the intellectual property (e.g., patents) is arranged in the consortium agreement.

3.2.3 Metadata

We will publish metadata along with all data published on the Zenodo platform. This ensures that the metadata is openly available and licenced under a public domain dedication CC0. The metadata will contain instructions to enable the user to access the data. The data and metadata will remain accessible and findable on the Zenodo platform after the end of the project. Data published on the Zenodo platform may require additional software to be accessible. The metadata will contain information on which software is required and if possible, references to open-source software needed to read the data will be provided.

3.3 Making data interoperable

Industry standardized file formats will be used to store benchmark data (e.g., .wav, mp4). Files will be converted to open file formats where possible for long-term storage. Whenever no industry standardized file format is available, customized solutions will be introduced. These may include text files or custom XML files with an associated schema to define its semantics. Documentation of these file formats will be included along with the files themselves to enable interoperability of the data contained in these files.

3.4 Increase data re-use

The generated data will be made comprehensible by incorporating readme files. Those readme files will explain in-depth the used methodology with respect to the process through which the

data was collected and validated. It will also specify the license under which the data can be used. The data, with accompanying metadata, will be shared in a timely manner. This timely release will be no later than the publication of the main findings and results and will be in line with established best practices in both research and in the specific field.

All non-confidential research data will be made available concurrently or shortly after publication by uploading it on the Zenodo platform. A licensing scheme will be established in accordance with the guidelines set in article 17(annex 5) of the Grant Agreement. Public research data will be retained for the lifetime of the Zenodo repository (but at least for 10 years after the end of the project).

More details regarding our policy to stimulate data re-use and licensing will be provided with future DMP updates.

4 Other research outputs

In addition to data, the CONVOLVE project will also generate other digital outputs (i.e., software, workflows, protocols, and models). Outputs that are marked by the consortium partners as public will be made available in the form of project reports and open access publications. These outputs will also be used in conference presentations and project-related workshops.

Furthermore, the Zenodo platform will be used to share any output marked as public domain. A special CONVOLVE community (zenodo.org/communities/convolve) has been created on this platform to make all CONVOLVE output easily findable and accessible. Appropriate licenses will be used on these artifacts to enable re-use, in line with the FAIR principles.

5 Allocation of resources

Lead for this data management task will be with TUE, though all partners are involved in the compliance of the DMP. The partners deliver datasets and metadata produced or collected in CONVOLVE according to the rules described in Section 3. The project coordinator will lead the implementation of the DMP and track the compliance of the rules as documented in this DMP. The CONVOLVE project partners have covered the costs for data preservation in their budget estimations. The Zenodo platform (zenodo.org/communities/convolve) will be used for long term preservation of public results. The choice of this platform ensures that no additional cost will be involved.

6 Data security

In this project several types of experimental and numerical data will be generated. The raw data will be stored by each partner according to their own standard procedures minimum for ten years after ending of the project. Furthermore, data sets and postprocessed data that needs to be accessed by multiple partners (but which is at that moment released to the public domain) will be stored on a local GitLab installation hosted by TUE (gitlab.tue.nl). Data within this repository is secured and backed-up according to the standard procedures used by the IT department of TU/e. Usage of this repository is convenient for sharing data in a secure manner. Long-term storage of public data is handled through our CONVOLVE community on Zenodo.

The research data used for communication, dissemination and exploitation will be stored on our internal Teams platform. This internal platform is only accessible for the project partners and has already been described in Deliverable D9.1.

Processed data will be made available to the community in three ways. First, the processed data will become available in the form of project reports and open access publications. Second, this data will be further exploited in webinars, articles in professional journals, and by conference presentations. Third, the Zenodo platform will be used for secure long-term storage and sharing of the processed data. A special CONVOLVE community (zenodo.org/communities/convolve) has been created on this platform to make all CONVOLVE data easily findable and accessible.

7 Ethics

There are no ethical or legal issues foreseen that can have an impact on data sharing. Since no personal data is being collected, no specific measures are needed with respect to informed consent for data sharing and long-term preservation of personal data?

8 Other issues

CONVOLVE will follow the DMP policies set by TUE. These policies are described at <https://www.tue.nl/en/our-university/library/library-for-researchers-and-phds/research-data-management/rdm-themes>