

D1.1 Data and Research Outputs Management Plan - a

Version 1.0

31 July 2023

Abstract

COGNIT is an AI-enabled Adaptive Serverless Framework for the Cognitive Cloud-Edge Continuum that enables the seamless, transparent, and trustworthy integration of data processing resources from providers and on-premises data centers in the cloud-edge continuum, and their automatic and intelligent adaptation to optimise where and how data is processed according to application requirements, changes in application demands and behaviour, and the operation of the infrastructure in terms of the main environmental sustainability metrics. This document provides detailed information about the Project's Data Management Plan (DMP) and how it is making its data/research outputs findable, accessible, interoperable, and reusable (FAIR).



Copyright © 2023 SovereignEdge.Cognit. All rights reserved.



This project is funded by the European Union's Horizon Europe research and innovation programme under Grant Agreement 101092711 – SovereignEdge.Cognit



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Deliverable Metadata

Project Title:	A Cognitive Serverless Framework for the Cloud-Edge Continuum
Project Acronym:	SovereignEdge.Cognit
Call:	HORIZON-CL4-2022-DATA-01-02
Grant Agreement:	101092711
WP number and Title:	WP1. Project Management
Nature:	R: Report
Dissemination Level:	PU: Public
Version:	1.0
Contractual Date of Delivery:	30/06/2023
Actual Date of Delivery:	31/07/2023
Lead Author:	Alberto P. Martí (OpenNebula)
Authors:	Michael Abdou (OpenNebula), Monowar Bhuyan (UMU), Ignacio M. Llorente (OpenNebula), Nikolaos Matskanis (CETIC), Paul Townend (UMU), Constantino Vázquez (OpenNebula).
Status:	Submitted

Document History

Version	Issue Date	Status ¹	Content and changes
0.1	27/07/2023	Draft	Initial Draft
0.2	28/07/2023	Peer-Reviewed	Reviewed Draft
1.0	31/07/2023	Submitted	Final Version

Peer Review History

Version	Peer Review Date	Reviewed By
0.1	28/07/2023	Monowar Bhuyan (UMU)
0.1	28/07/2023	Nikolaos Matskanis (CETIC)

Summary of Changes from Previous Versions

First Version of the "Data and Research Outputs Management Plan" Deliverable

¹ A deliverable can be in one of these stages: Draft, Peer-Reviewed, Submitted, and Approved.

Executive Summary

This is the first version of Deliverable D1.1, the Data and Research Outputs Management Plan in WP1 ("Project Management"). It describes how data management will be performed by the COGNIT Project according to FAIR principles (Findable, Accessible, Interoperable, Reusable) and in adherence to the GDPR and other relevant regulation.

It is the intention of the Consortium to make all the data and research outputs openly available, following the principle "as open as possible, as closed as necessary". As a general rule, the Project will make its publicly available data and research outputs findable through **Zenodo**, given that this well-established platform follows the FAIR principles and is OpenAIRE-compliant. At a minimum, all scientific publications will be available via Green Open Access. All metadata may be freely used under Creative Commons licence CC0.

The project website will list all its publicly available data and research outputs, including links to the project deliverables and scientific publications on Zenodo, to other materials hosted locally or somewhere else (e.g. blog posts, press releases, commercial publications, screencasts, and webinars), as well as to the Project's **GitHub** public repository. A Creative Commons licence will be used for the project website and its associated online contents.

A central aim of the Project is to provide value to as many European organisations and citizens as possible beyond the original Consortium. By default, all software produced by COGNIT will be released as **open source** under Apache License, Version 2.0. All the source code produced by the Project, as well as eventually all AI/ML models (and the relevant metrics used to train them) and the datasets produced during the validation phases of the COGNIT Framework by its Use Cases, will be made available through the existing GitHub public repository. Apart from the secure use of Zenodo and GitHub repositories, personal and research data collected during the execution of the Project will be stored and managed centrally through OpenNebula's corporate account on **Google Workspace**.

The execution and regular updating of this Data and Research Outputs Management Plan will be performed by the Project Coordinator (OpenNebula) as Task Leader of T1.5 ("Data, Legal, Gender, and Ethical Framework Definition"). As such, OpenNebula has defined the figure of a **Data Management Officer (DMO)**, who will be in charge of overseeing the successful application of this Plan. OpenNebula has also appointed a **Data Protection Officer (DPO)** for the purposes of this project.

This deliverable covers up until the end of the first semester of the Project (M6), and will be revised towards the end of the Project's lifetime in months M18 and M36.

Table of Contents

Abbreviations and Acronyms	5
1. Introduction	6
2. Data Summary	7
3. FAIR data	9
3.1. Making data findable, including provisions for metadata	9
3.2. Making data accessible	10
3.3. Making data interoperable	12
3.4. Increase data re-use	14
4. Other research outputs	16
5. Allocation of resources	17
6. Data security	18
7. Ethics	20
8. Other issues	21
Annex 1. Open Science provisions in the GA	22

Abbreviations and Acronyms

AI	Artificial Intelligence
API	Application Programming Interface
CDPA	Cloud Data Processing Addendum
CLA	Contributor Licence Agreement
DMO	Data Management Officer
DOI	Digital Object Identifier
DPM	Data Management Plan
DPO	Data Protection Officer
GA	Grant Agreement
GDPR	General Data Protection Regulation
HTTP	Hypertext Transfer Protocol
JSON	JavaScript Object Notation
ML	Machine Learning
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
REST	Representational State Transfer
SQL	Structured Query Language
URL	Uniform Resource Locator

1. Introduction

The general purpose of Deliverable D1.1 is to describe in detail an extended version of a Data Management Plan, and to explain how the COGNIT Project is planning to make its data and research outputs findable, accessible, interoperable, and reusable, in line with the FAIR principles and their application to Horizon Europe projects (see Annex 1). In order to keep this plan up-to-date, D1.1 will be revised in M18 and in M36, incorporating any significant changes to the Project's Data Management policy that this document defines.

The structure of Deliverable D1.1 follows version 1.0 of the official Data Management Plan Template for Horizon Europe published by the European Commission on May 5, 2021. D1.1 is composed of seven main content chapters, including a Data Summary (Section 2), FAIR Data (Section 3), Other Research Outputs (Section 4), Allocation of Resources (Section 5), Data Security (Section 6), Ethics (Section 7), and Other Issues (Section 8). The report ends with an Annex that reflects the Open Science provisions in the COGNIT Grant Agreement.

This document includes references to the modular architecture of the COGNIT Framework (defined in Deliverable D2.1) and to the planned actions to communicate and disseminate the Project's results among the scientific community (detailed in Deliverable D6.1).

2. Data Summary

The Project is expected to generate and process a variety of data for the validation of the research experiments reflected in the publications to be produced during its execution. All generated data will be stored in an easily accessible way by both humans and software, and grouped in two main categories: (i) personal data, collected as part of the Community and Dissemination actions in WP6, and (ii) research data, including experimental data, scientific publications, and software. Whereas personal data will not be made available, in line with the GDPR, research data might come in different forms, namely:

- **Project deliverables** – Published as part of the requirements defined by the COGNIT Grant Agreement but also as one of the main channels to share the results and research data of the Project. Once submitted to the EC, all the Deliverables (unless defined as “Sensitive”) will be made available as PDF files on the project website and, once approved by the EC, also on Zenodo.
- **Scientific publications** – At a minimum, all scientific publications produced by COGNIT will be available via Green Open Access. The Consortium Agreement, signed by all partners in the Project, defines the process by which results of the Project will be disseminated as academic papers. The Scientific Coordinator of the Project (UMU) will oversee that partners abide by those rules and also to the data management policy defined in this document.
- **Other publications and materials** – Beyond scientific publications, COGNIT is expected as part of its Communication and Dissemination Plan (see Deliverable D6.1) to produce and share additional data, such as:
 - Blog posts on the project website.
 - Official press releases.
 - Reports describing the contribution to standards.
 - Commercial publications in online magazines and third-party blogs.
 - Screencasts and video tutorials.
 - Promotional materials for events (e.g. flyers and roll-ups)
- **Software** – By default, all software produced by COGNIT will be released as open source under Apache License, Version 2.0. Eventual upstream contributions to relevant European open source technologies will be done in accordance with their existing open source licence and CLA. The Project will host and share all the source code, including the implementation of REST APIs, modules, and other software artefacts, in a GitHub public repository.
- **Metrics and AI/ML models** – COGNIT involves the use of Artificial Intelligence for optimising the orchestration and management of large-scale edge deployments using infrastructure resources in multi-provider cloud environments. The final goal is to develop a sustainable smart edge platform offering placement algorithms to

minimise the environmental impact of deploying applications at the edge. For that purpose, the Project will produce AI/ML models and a body of experimental data based on the collection of metrics generated by the COGNIT Framework and used to train and improve those models (see D2.1 for more details).

- **Use Case datasets** – Some of the Use Cases cases involved in the validation of the COGNIT Framework (see Deliverable D5.1 for more details) might produce their own experimental data during their agile phases of research and development. These datasets may include simulations of real-life latency scenarios, specific observability data (e.g. logs, metrics, traces, and events), or the conclusions of anomaly detection mechanisms involved in cybersecurity analyses.

The data generated within the Project will be traceable and locatable by means of unique identification mechanisms. All generated research data will be properly annotated with metadata, and datasets will be shared through the Project's Zenodo repository.

Data will be accessible between partners through a common space for the necessary data exchange and communication; certain sensitive research data produced by Use Cases may fall under GDPR protection or might not be made publicly available due to IPR restrictions or potential exploitation usage by its owners. Future versions of this document will report if such a situation finally takes place.

Data will be made available as soon as possible and published on the project website and other relevant repositories (e.g. Zenodo or GitHub). Access will be permitted to anyone, in order to use a data copy which may be shared or adapted (transformed or built upon existing material). No embargoes are in place at the time of writing this report.

OpenNebula, as Project Coordinator, is responsible for maintaining the Project's public data repositories and the internal collaborative tools needed for the implementation of the Project. Each partner is responsible for the recoverability of their own generated data. The costs of making data FAIR cannot be determined yet, as it depends on the amount of data which will be generated and the cost of the long-term storage solution. The costs of Open Access publications have been included in the project budget of most of the partners of COGNIT.

Given that this first version of the Data and Research Outputs Management Plan covers up until M6, a number of details regarding the Project's research data cannot be specified yet. Future versions of this document will incorporate that information, including the origin of each dataset, the applicable IPR, their size and format, and their relevance to third-parties.

3. FAIR data

3.1. Making data findable, including provisions for metadata

As a general rule, the COGNIT Project will make its publicly available data and research outputs findable through **Zenodo**, given that this well-established platform follows the FAIR principles and is OpenAIRE-compliant:

- ✓ (Meta)data is assigned a globally unique and persistent identifier: A Digital Object Identifier (DOI) is issued to every published record on Zenodo.
- ✓ Data are described with rich metadata: Zenodo's metadata is compliant with DataCite's Metadata Schema minimum and recommended terms, with a few additional enrichments.
- ✓ Metadata clearly and explicitly include the identifier of the data it describes: The DOI is a top-level and a mandatory field in the metadata of each record.
- ✓ (Meta)data are registered or indexed in a searchable resource: Metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing. Metadata of each record is sent to DataCite servers during DOI registration and indexed there. The following metadata formats are provided by Zenodo: MARCXML, Dublin Core (according to OpenAIRE Guidelines), DataCite, DCAT, JSON-LD (Schema.org). Keywords are provided in the metadata to optimise the possibility for discovery and re-use.

The data generated by the Project will be traceable and locatable by means of a unique DOI. A DOI is a character string used to uniquely identify an object such as an electronic document. Metadata about the object is stored in association with the DOI name and this metadata may include a location, such as a URL, where the object can be found. The DOI for a document remains fixed over the lifetime of the document, whereas its location and other metadata may change. Referring to an online document by its DOI provides more stable linking than simply referring to it by its URL. If its URL changes, the metadata for the DOI can be updated so that it links to the new URL.

Scientific papers produced out of the Project's research results are expected to have DOIs assigned by the editorial companies. In any case Zenodo also assigns all publicly available uploads a DOI to make the upload easily and uniquely citable. Zenodo registers DOIs automatically for all deposited materials. Zenodo supports harvesting of metadata via the OAI-PMH protocol.



Zenodo allows users to create collections and accept or reject uploads submitted to it. By the end of M7, the Project will create a public collection for listing its data: zenodo.org/communities/sovereignedgeeu-cognit

Several of the researchers involved in the COGNIT Project have personal pages on scientific social networking sites (e.g. ResearchGate, Academia.edu, LinkedIn). These platforms provide an additional channel to disseminate the Project's results and make its publicly available data and research outputs findable. As described in Deliverable D6.1, the Project will undertake several specific actions to communicate and disseminate its results among the scientific community.

The project website (COGNIT.SovereignEdge.eu)—currently based on a self-managed WordPress instance—will list all its publicly available data and research outputs, including links to the project deliverables and scientific publications on Zenodo, to other materials hosted locally or somewhere else (e.g. blog posts, press releases, commercial publications, screencasts, and webinars), as well as to the Project's **GitHub** public repository, leveraging the fact that Zenodo has a native integration with GitHub to make code and other resources on GitHub citable².

3.2. Making data accessible

It is the intention of the Consortium to make all the data and research outputs openly available, following the principle “as open as possible, as closed as necessary”.

Repository:

The COGNIT Project will deposit its publicly available data and research outputs in open online repositories. For project deliverables and scientific publications, the primary location will be the universal Zenodo repository, whose main features in terms of providing a unique identifier to digital objects have already been described above. UMU's institutional repository (based on Sweden's DiVA) might also be used for deposition of the publications produced by the Project, in line with the University's open access policy.³

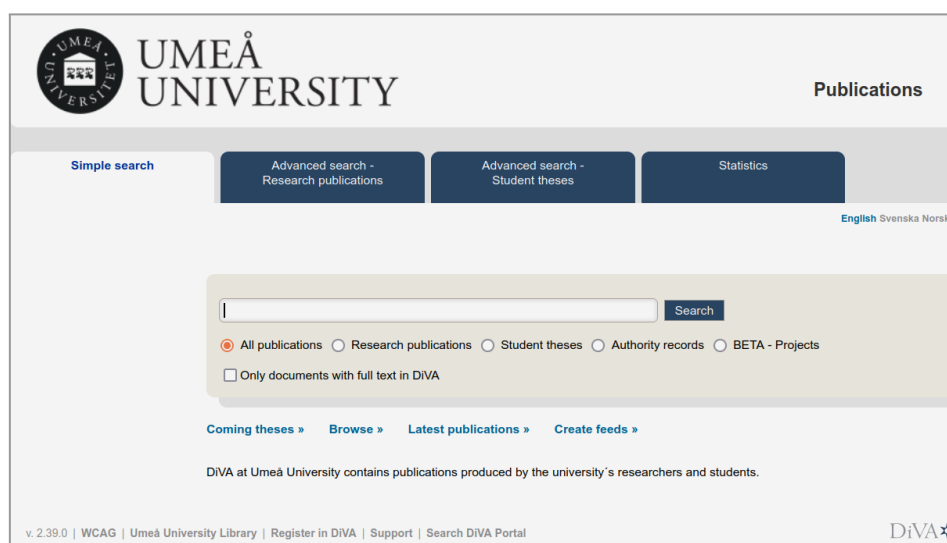


Figure 3.1. Interface of UMU's academic repository.

² <https://docs.github.com/en/repositories/archiving-a-github-repository/referencing-and-citing-content>

³ <https://www.umu.se/en/library/publish/open-access/>

All the source code produced by the Project, as well as eventually the AI/ML models (and the relevant metrics used to train them) and the datasets produced during the validation phases of the COGNIT Framework by its Use Cases, will be made available through the existing GitHub public repository: github.com/SovereignEdgeEU-COGNIT

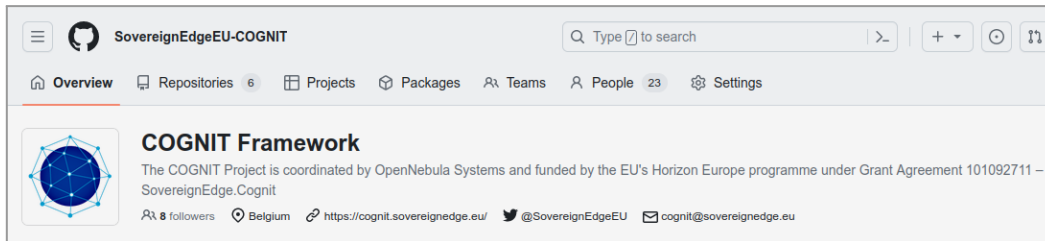


Figure 3.2. The Project's existing GitHub public repository.

Data:

The list of project deliverables catalogued as "Public" in the GA, for example, will be available as PDF files from the Resources section of the project website just after its submission to the European Commission. These provisional documents will be updated once submitted deliverables are finally approved by the EC, and uploaded to Zenodo. Project deliverables defined as "Sensitive" in the GA, however, will not be publicly available and will be kept only for distribution within the Consortium:

Number	Name	Work Package	Due Date
D6.4	Exploitation Plan - a	WP6	M18
D6.5	Exploitation Plan - b	WP6	M36

Table 3.1. List of project deliverables catalogued as "Sensitive" in the Grant Agreement.

At a minimum, all scientific publications produced during the execution of the Project will be available via Green Open Access. Consortium partners will be encouraged to publish via Gold Open Access, using in-kind contributions from their institutions to fund this and only if necessary. No embargoes are in place at the time of writing this report.

A central aim of the Project is to provide value to as many European organisations and citizens as possible beyond the original Consortium. By default, all software produced by COGNIT will be released as open source under [Apache License, Version 2.0](#). Eventual upstream contributions to relevant European open source technologies (e.g. OpenNebula, SUSE's OPNI, or Phoenix-RTOS) will be done in accordance with the existing open source licence and Contributor License Agreement (CLA) defined by each project.

A Creative Commons licence similar to the one applied to public deliverables (i.e. [CC BY-NC-SA 4.0](#)) will be used for the project website and other online contents (e.g. blog posts and promotional materials used in Communication and Dissemination actions).

Some data produced during the execution of the Project (e.g. AI/ML models, the metrics used to train them, or specific Use Case datasets) might not be made publicly available due

to IPR restrictions or potential exploitation usage by its owners (i.e. members of the Consortium, individually or jointly). At the time of writing this Deliverable D1.1, it is too early to confirm these intentions or anticipate specific legal and contractual reasons for those potential restrictions. If that was the case, a full report will be provided in future updates of this Data and Research Outputs Management Plan in either M18 or M36.

Personal data collected during the execution of the Project as part of its Communication and Dissemination actions (e.g. event registrations) will follow the “data minimisation principle” and relevant regulations (e.g. GDPR), and will only be available internally to those members of the Consortium directly involved in the tasks carried out by the Project’s WP6 (“Dissemination, Communication, Exploitation, and Standardization”).

Metadata:

The Project will leverage the standard features of Zenodo regarding metadata:

- ✓ All metadata may be freely used under Creative Commons licence [CC0](#).
- ✓ (Meta)data are retrievable by their identifier using a standardised communications protocol: Metadata for individual records as well as record collections are harvestable using the OAI-PMH protocol by the record identifier and the collection name. Metadata is also retrievable through the public REST API.
- ✓ The protocol is open, free, and universally implementable: OAI-PMH and REST are open, free and universal protocols for information retrieval on the web.
- ✓ The protocol allows for an authentication and authorization procedure, where necessary: Metadata are publicly accessible and licensed under public domain. No authorization is ever necessary to retrieve it.
- ✓ Metadata are accessible, even when the data are no longer available: Data and metadata will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least. Metadata is stored in high-availability database servers at CERN, which are separate from the data itself.

3.3. Making data interoperable

For all the resources hosted on Zenodo, the Project will leverage its standard features regarding metadata:

- ✓ (Meta)data uses a formal, accessible, shared, and broadly applicable language for knowledge representation: Zenodo uses JSON Schema as internal representation of metadata and offers export to other popular formats such as Dublin Core or MARCXML.
- ✓ (Meta)data uses vocabularies that follow FAIR principles: For certain terms we refer to open, external vocabularies, e.g.: licence (Open Definition), funders (FundRef) and grants (OpenAIRE).

- ✓ (Meta)data include qualified references to other (meta)data: Each referenced external piece of metadata is qualified by a resolvable URL.

The implementation by the Project of REST APIs, modules, and other software artefacts as open source software will follow GitHub's best practices for projects⁴ and repositories.⁵ The JSON format will be the standard format for data exchange between the different components of the COGNIT Framework (see Deliverable D2.1 for more details).

```
"HOST": {
  "ID": "1",
  "NAME": "edge-cluster-host-1",
  "STATE": "2",
  "PREV_STATE": "2",
  "IM_MAD": "kvm",
  "VM_MAD": "kvm",
  "CLUSTER_ID": "0",
  "CLUSTER": "default",
  "HOST_SHARE": {
    "MEM_USAGE": "196722688",
    "CPU_USAGE": "7100",
    "TOTAL_MEM": "792421016",
    "TOTAL_CPU": "12800",
    "MAX_MEM": "792421016",
    "MAX_CPU": "12800",
    "RUNNING_VMS": "101",
    "VMS_THREAD": "1",
    "DATASTORES": {
      "DISK_USAGE": "0",
      "FREE_DISK": "5682015",
      "MAX_DISK": "12015252",
      "USED_DISK": "5842850"
    },
    "PCI_DEVICES": { (...) },
    "NUMA_NODES": { (...) }
  },
  "VMS": { (...) },
  "TEMPLATE": {
    "ARCH": "x86_64",
    "CGROUPS_VERSION": "2",
    "CPUSPEED": "0",
    "HOSTNAME": "supermicro3",
    "HYPERVISOR": "kvm",
    (...)
  },
  "MONITORING": { (...) }
}
```

Table 3.2. Fragment of a JSON file describing a virtualisation host.

⁴ <https://docs.github.com/en/issues/planning-and-tracking-with-projects/learning-about-projects/best-practices-for-projects>

⁵ <https://docs.github.com/en/repositories/creating-and-managing-repositories/best-practices-for-repositories>

Data interoperability will also be based on the consistent use during the execution of the Project of a series of recommended formats:

Type	Recommended formats
Project deliverables	PDF
Other textual documents	TXT, ODT, XML, HTML, PDF
Images	JPEG, GIF, PNG
Audio	MP3
Video	MP4
Databases	CSV, ODB, SQL
Software	Bash, C/C++, JSON, Python, YAML

Table 3.3. Initial set of recommended formats per each type of data used in the Project.

This initial set of recommended formats will be updated in future versions of this report, incorporating changes derived specially from the research carried out by the Use Cases.

3.4. Increase data re-use

Data and research outputs produced during the execution of the Project will be made available to all third-parties as soon as possible and published on the project website and on other relevant repositories. Access will be permitted to anyone, also after the end of the Project, in order to use a data copy which may be shared or adapted (transformed or built upon existing material), always making sure that a proper attribution to COGNIT has been included. The project website will be accessible for 5 years after the end of the Project, whereas the data hosted on the Zenodo and GitHub repositories is expected to stay available indefinitely.

For the source code and other resources hosted on GitHub, reusability will be supported by the fact that all software produced by COGNIT will be released as open source under [Apache License, Version 2.0](#), as well as by the standard practice of including a detailed README file in each section.⁶ No data embargoes are in place at the time of writing this report. Should that happen in the future, revised versions of this document will include all the necessary details about the embargoes being adopted after M6.

The production of scientific papers will be one of the main means of public disclosure of research data produced by the Project. All researchers involved in these publications will ensure that the underlying data will be made available for verification purposes after publication. After publication, and assuming no temporary embargo is imposed by the

⁶ <https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/about-readmes>

publisher, the underlying data relevant to those scientific papers will be immediately uploaded into the Project's public repositories—if not present already.

For all the resources hosted on Zenodo, the Project will leverage its standard features regarding metadata:

- ✓ (Meta)data are richly described with a plurality of accurate and relevant attributes: Each record contains a minimum of DataCite's mandatory terms, with optionally additional DataCite recommended terms and Zenodo's enrichments.
- ✓ (Meta)data are released with a clear and accessible data usage licence: Licence is one of the mandatory terms in Zenodo's metadata, and is referring to an Open Definition licence. Data downloaded by the users is subject to the licence specified in the metadata by the uploader.
- ✓ (Meta)data is associated with detailed provenance: All data and metadata uploaded is traceable to a registered Zenodo user. Metadata can optionally describe the original authors of the published work.
- ✓ (Meta)data meet domain-relevant community standards: Zenodo is not a domain-specific repository, yet through compliance with DataCite's Metadata Schema, metadata meets one of the broadest cross-domain standards available.

4. Other research outputs

As mentioned in the previous chapter, at the time of writing this Deliverable D1.1 (i.e. M6) it is not possible to confirm to what extent the FAIR principles will end up being applied to the AI/ML models, the relevant metrics used to train them, and the eventual Use Case datasets that might be produced during the execution of the Project, and whether they might or might not be made publicly available due to IPR restrictions or potential exploitation usage by its owners. In any case, a full report about those research outputs will be provided in future updates of this Data and Research Outputs Management Plan.

5. Allocation of resources

The Project Coordinator is responsible for setting up and maintaining the Project’s repositories (i.e. Zenodo and GitHub)—which are expected to remain free-of-charge in the future—as well as the infrastructure that will support the project website up until 5 years after the end of the Project. Beyond the role that the Project’s Scientific Advisory Board (SAB) is expected to play after its establishment in M7 in terms of evaluating research objectives and outcomes, the Scientific Coordinator (UMU) remains ultimately responsible for the quality assurance related to scientific data and research outcomes. On the other hand, the Technical Manager (OpenNebula) is responsible for the quality assurance related to any technical/software outcomes, using as a reference for the whole Consortium the existing Quality Assurance Policy of OpenNebula when it comes to open source software development performed by the COGNIT Project.⁷

The execution and regular updating of this Data and Research Outputs Management Plan will be performed by the Project Coordinator (OpenNebula) as Task Leader of T1.5 (“Data, Legal, Gender, and Ethical Framework Definition”). As such, OpenNebula has defined the figure of **Data Management Officer (DMO)**, who will be in charge of overseeing the successful application of the Data and Research Outputs Management Plan during the execution of the Project and coordinating the work carried out by the Scientific Coordinator and the Technical Manager. The contact details of the current DMO are:

Name	Dr. Alberto P. Martí
Company	OpenNebula Systems
Address	La Finca Business Park, Building #4 28223 Pozuelo de Alarcón (Spain)
Telephone	+34918298445
E-mail	innovation@opennebula.io

Table 5.1. Contact details of the current PDO of the Project.

Despite not being required under the GDPR to do so, OpenNebula as Project Coordinator has also appointed a **Data Protection Officer (DPO)** for the purposes of this project. The contact details of the DPO—both an email (dpo@opennebula.io) and a postal address—will be available to all data subjects somehow involved in COGNIT.

⁷ <https://github.com/OpenNebula/one/wiki/Quality-Assurance>

6. Data security

Apart from the secure use of Zenodo and GitHub repositories, personal and research data collected during the execution of the Project will be stored and managed centrally through OpenNebula's corporate account on **Google Workspace**—Google's cloud productivity tool. In line with OpenNebula's current policies, appropriate measures have been put in place for securing all data managed by OpenNebula as Project Coordinator. Similar data security recommendations have been circulated among the rest of partners in the COGNIT Project:

- Access to company premises and office spaces, with special measures being applied to rooms hosting corporate IT infrastructure.
- Correct disposal of paperwork and devices, along with protections for those that have been accidentally lost.
- Integrity and cyber-protection of all data, applications, corporate systems and networks, including use of firewalls, malware scans, anti-virus protection and regular installation of security patches and updates.
- Implementation of a strong password policy and secure remote access.
- Use of encryption and/or pseudonymisation where appropriate to do so.
- Access to shared resources hosted by public cloud/edge providers.

OpenNebula will require all third-party service providers (i.e. Zenodo, GitHub, and Google) to respect the security of the personal and research data to be collected during the Project, and to treat it in accordance with the GDPR regulatory framework and standard data security principles. Third-party service providers shall not be allowed to process this data for their own purposes or for purposes other than the ones described in this report. The Consortium's access to its shared Google Workspace, for instance, is covered by the fact that OpenNebula has signed Google's Cloud Data Processing Addendum (CDPA),⁸ which incorporates Standard Contract Clauses as a means of meeting the security, contracting, and data transfer requirements under EU data protection regulations.

Given that Communication and Disseminations tasks in WP6 require direct interaction with external users (e.g. through attendance to webinars or physical workshops), a minimum of personal data needs to be collected in order to manage the execution of those tasks and maintain a list of contact persons interested in COGNIT so that an effective communication channel with them can be maintained during the execution of the project. The personal data collected for this purpose will be restricted to:

- Full name
- Organisation
- Job title
- Corporate email address

⁸ <https://cloud.google.com/terms/data-processing-addendum>

Personal data collection will be performed through secure, on-line self-registration forms and following a GDPR-compliant process. Access to data will be encrypted and protected by authorisation and authentication mechanisms. The data will not be retained beyond the duration of the project.

In order to ensure appropriate security of the personal data (including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage), and in line with OpenNebula's current policies, our approach to safeguard the data involved in this project includes the following principles and measures:

- Undertaking regular analyses of the risks presented by our processing of research and personal data, and use this to assess the appropriate level of security we need to put in place at each phase of the project.
- Making sure that we regularly review our information security policies and measures and, where necessary, improve them.
- We understand the requirements of confidentiality, integrity and availability for the personal data we process.
- Thanks to an appropriate backup process, we guarantee business continuity and make sure that we can restore access to research and personal data in the event of any incidents.
- We conduct regular testing and reviews of our measures to ensure they remain effective, and act on the results of those tests where they highlight areas for improvement.
- When using the services of a third-party data processor (including those commercialised by entities in non-EU countries), we make sure they also implement appropriate, GDPR-compliant technical and organisational measures.
- Our DPO is actively involved in educating the company, its subsidiaries, and employees on GDPR compliance and relevant data security best practices.

As a general rule, each partner in COGNIT is responsible for the recoverability of their own generated data.

7. Ethics

Questionnaires dealing with personal data (e.g. as part of Communication & Dissemination actions) will be GDPR-compliant and shall not require informed consent for data sharing and long term preservation. As a general rule, the Project shall abide by the principles relating to processing of personal data defined in Article 5 of the GDPR, meaning that all personal data managed during the execution of the COGNIT Project shall be:

- Processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
- Collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');
- Adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');
- Accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');
- Kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');
- Processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

As stated at the original Ethics Self-Assessment, the COGNIT Project involves the use of Artificial Intelligence for optimising the orchestration and management of large-scale edge deployments using infrastructure resources in multi-provider cloud environments. This approach enables deployments across the Datacenter-Cloud-Edge-IoT continuum to act proactively with a view to anticipating problems, increasing security and reducing consumption of resources. The final goal is to develop a sustainable smart edge platform offering placement algorithms to minimise the environmental impact of deploying applications at the edge. In that sense, it is our understanding that the use of AI in this Horizon Europe project does not raise additional ethical issues nor has any impact on the data sharing policy of the Project as defined in this Data and Research Outputs Management Plan. Should that situation change during the course of the Project, future versions of this document must then include a full assessment of that potential impact.

8. Other issues

No other national/funder/sectorial/departmental procedures for data management are expected to be used during the execution of the Project. Future versions of this document will reflect any changes in that regard.

Annex 1. Open Science provisions in the GA

Open science: open access to scientific publications

The beneficiaries must ensure open access to peer-reviewed scientific publications relating to their results. In particular, they must ensure that:

- *at the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications*
- *immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights; for monographs and other long-text formats, the licence may exclude commercial uses and derivative works (e.g. CC BY-NC, CC BY-ND) and*
- *information is given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication.*

Beneficiaries (or authors) must retain sufficient intellectual property rights to comply with the open access requirements.

Metadata of deposited publications must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: publication (author(s), title, date of publication, publication venue); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the publication, the authors involved in the action and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication.

Only publication fees in full open access venues for peer-reviewed scientific publications are eligible for reimbursement.

Open science: research data management

The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions:

- *establish a data management plan ('DMP') (and regularly update it)*
- *as soon as possible and within the deadlines set out in the DMP, deposit the data in a trusted repository; if required in the call conditions, this repository must be federated in the EOSC in compliance with EOSC requirements*
- *as soon as possible and within the deadlines set out in the DMP, ensure open access — via the repository — to the deposited data, under the latest available version of the*

Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence with equivalent rights, following the principle 'as open as possible as closed as necessary', unless providing open access would in particular:

- *be against the beneficiary's legitimate interests, including regarding commercial exploitation, or*
- *be contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations under this Agreement; if open access is not provided (to some or all data), this must be justified in the DMP*
- *provide information via the repository about any research output or any other tools and instruments needed to re-use or validate the data.*

Metadata of deposited data must be open under a Creative Commons Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: datasets (description, date of deposit, author(s), venue and embargo); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for related publications and other research outputs.

Open science: additional practices

Where the call conditions impose additional obligations regarding open science practices, the beneficiaries must also comply with those.

Where the call conditions impose additional obligations regarding the validation of scientific publications, the beneficiaries must provide (digital or physical) access to data or other results needed for validation of the conclusions of scientific publications, to the extent that their legitimate interests or constraints are safeguarded (and unless they already provided the (open) access at publication).

Where the call conditions impose additional open science obligations in case of a public emergency, the beneficiaries must (if requested by the granting authority) immediately deposit any research output in a repository and provide open access to it under a CC BY licence, a Public Domain Dedication (CC 0) or equivalent. As an exception, if the access would be against the beneficiaries' legitimate interests, the beneficiaries must grant non-exclusive licenses — under fair and reasonable conditions — to legal entities that need the research output to address the public emergency and commit to rapidly and broadly exploit the resulting products and services at fair and reasonable conditions. This provision applies up to four years after the end of the action (see Data Sheet, Point 1).