



Funded by
the European Union

Grant Agreement number: 101092696

Topic: HORIZON-CL4-2022-DATA-01



CODECO

Cognitive Decentralised
Edge Cloud Orchestration

D2: Data Management and Ethics Handling v1.0

Work package	WP1 – Project management, scientific coordination, and data management
Deliverable Internal Number	D1.2
Task	Task 1.1
Due date	31.03.2023
Submission date	31.03.2023
Dissemination Type	Public
Deliverable lead and editor	INO, Vitor Vieira
Contributing Partners	All partners
Version	1.0
Reviewer 1	FOR, Rute C. Sofia
Reviewer 2	INTRA, John Soldatos



Funded by
the European Union

Project Partners



Funded by
the European Union

Executive Summary

D2 is a deliverable of WP1 which covers the *Data Management and Ethics Handling* as a key element of good data management in Horizon Europe projects. As part of making research data findable, accessible, interoperable, and re-usable (FAIR), we also consider the ethical dimensions of handling the IoT data that is necessary for the use-cases, and how to correctly manage them throughout the project.

Keywords: Cloud, Data Collection, Data Management, Edge, Ethics, Internet of Things, Internetworking, Network Technologies, Open Source, Software.

Document Revision History

Version	Date	Description of change	List of contributor(s)
v0.1	02.03.2023	Initial draft	Vitor Vieira, INO
v0.2	29.03.2023	Internal review by FOR	Rute C. Sofia, FOR
v0.3	30.03.2023	Addition of inputs after revision by FOR	Vitor Vieira, INO Joana Rodrigues, INO
v0.4	30.03.2023	Second revision by FOR	Rute C. Sofia, FOR
v0.5	31.03.2023	Quality Control and revision by INTRA	John Soldatos, INTRA
v0.6	31.03.2023	Revision by FOR	Rute C. Sofia, FOR
v0.7	31.03.2023	Integration of reviews by INO	Vitor Vieira, INO
v0.8	31.03.2023	Release to the Coordinator	Vitor Vieira, INO
v1.0	31.03.2023	Final verification by the Coordinator and release to the EC	Rute C. Sofia, FOR

Disclaimer

The information, documentation and figures available in this deliverable have been developed by the Horizon Europe CODECO project consortium, under the European Union grant Agreement number 101092696. The content does not necessarily reflect the views of the European Commission. The European Commission is not liable for any use that may be made of the information contained herein.

Copyright notice: © 2023 – 2025 CODECO Consortium



Funded by
the European Union

Table of Contents

Executive Summary	3
1 Introduction	8
1.1 Document Structure	9
1.2 Dependencies	10
2 Data Summary	10
2.1 Types of Data	10
2.2 Data Collection	11
2.3 Data Formats	12
2.4 Template for Collecting Datasets	12
3 FAIR Data	13
3.1 Making Data Findable and Openly Accessible, Including Provisions for Metadata	13
3.2 Making Data Interoperable	15
3.3 Increase Data Reuse	15
3.4 Initial Dataset Handling	15
4 Allocation of Resources	16
4.1 Responsible Entities	16
4.2 Responsible Persons	16
4.3 Curation	16
5 Data Preservation, Storage and Privacy	17
5.1 Data Storage and Backup	17
5.2 Data Selection and Preservation	17
5.3 Data Destruction	18
6 Ethical Aspects	18



6.1	Processes	19
6.2	GDPR Compliance.....	20
7	Other Issues	20
8	Use-cases Initial Considerations	20
8.1	Smart Cities	20
8.2	Energy	21
8.3	Manufacturing	21
9	Summary	22
	Annex 1 – Dataset Information Collection Template.....	23

List of Tables

Table 1: Categories of datasets expected in CODECO.	11
Table 2: Initial data formats used in CODECO.	12
Table 3: Responsible Entities and Persons.	16

List of Acronyms

Acronym	Meaning
3GPP	3 rd Generation Partnership Project
AAA	Authentication, Authorization, Accounting
AI	Artificial Intelligence
CA	Consortium Agreement
CC	Creative Commons
CDN	Content Delivery Network
CLI	Command Line Interface
CODECO	Cognitive, Decentralised Edge-Cloud Orchestration
CSV	Comma Separated Values
DL	Deep Learning
DME	Data Management and Ethics Handling
DMP	Data Management Plan
DOI	Digital Object Identifier
EC	European Commission



Acronym	Meaning
EN	Edge Node
ETC	Ethics Committee
FAIR	Findable, Accessible, Interoperable, and Re-Usable
HE	Horizon Europe
IoT	Internet of Things
IP	Internet Protocol
JSON	JavaScript Object Notation
K8s	Kubernetes
PDF	Portable Data Format
PDLC	Privacy-preserving Decentralised Learning and Context-awareness
PII	Personally Identifiable Information
QoE	Quality of Experience
QoS	Quality of Service
WP	Work Package
WPL	Work Package Leader



Acknowledgements

We would like to thank all the partners that contributed to the development of the CODECO DME. Special thanks to the Project Coordinator Rute Sofia from FOR.



**Funded by
the European Union**

1 Introduction

CODECO Deliverable D2 - *Data Management and Ethics Handling (DME)* is a document detailing the management of all the data generated by the project, developed in the context of WP1, T1.1 – Project and Data Management. The aim of this deliverable is to provide data management guidelines following the FAIR (Findable, Accessible, Interoperable, Reusable) principles.

The DME establishes guidelines towards the handling of data, protection of personal, business, and confidential data. DME provides an initial overview on dataset formats in CODECO, and how such datasets will be provided to the public, to the consortium. Instructions for storage, generation and deletion are also detailed.

The procedures established aim at opening as much as possible the generated datasets, based on open repositories, namely, the CODECO Zenodo community, the Website, and the Eclipse (ECL) CODECO research lab. Datasets shall be identified by a *Digital Object Identifier (DOI)* and follow as much as possible open licensing.

CODECO is an orchestration framework that provides support for data management, decentralised data workflow, dynamic computation offloading, adaptive networking services, and privacy-preserving decentralized learning mechanisms. It includes open cognitive toolkits, a developer-oriented open-source software repository, training tools, use-cases across multiple domains, open calls, and community events, and integration into EdgeNet for experimentation. In this context the data related to the use-cases and project development, even more the collect IoT data is of extreme relevance to the DMP.

As a research project, CODECO is concerned with proper and effective management of reach data, which is a very broad term that comprises different types of data. In this direction, the project adheres to best practices regarding:

- (i) The preparation of DME as a living document that will be regularly updated throughout the project's lifetime.
- (ii) The transfer and deposit of data in a trusted repository such as ZENODO, which will be openly accessible following the principle: "*as open as possible, as closed as necessary*".
- (iii) The provision of information about any research output and other tools that can be used to repurpose, reuse, and validate the data.



DME is therefore a key element of good data management in Horizon Europe projects. As part of making research data findable, accessible, interoperable, and re-usable (FAIR), the CODECO's DME sets the following parameters:

- What data will be collected per use-case, processed, and/or generated.
- The handling of research data during and after the end of the project.
- Which methodology will be applied.
- Whether data will be shared/made open access for verification and reuse and how.
- How data will be curated and preserved during and after the end of the project.
- How the data will preserve privacy and be secure throughout the project.

The DME is intended to be a “living document” in which information can be made available with detail through updates as the implementation of the project progresses and when significant changes occur, such as the production of new datasets, changes in data management infrastructures (e.g., open access repositories), changes in data management policies (e.g., confidentiality levels of certain datasets), changes in consortium policies (e.g., new innovation potential, the decision to file for a patent), changes in consortium composition and external factors (e.g., new consortium members joining or old members leaving).

A first important step to updating this living document will be the collection of information about the main datasets that will be produced and managed in the project. This collection will be performed based on the template that is presented in Appendix 1. A first collection shall be completed until month 4 (April 2023) of the project, and WP1 shall oversee the regular update of the datasets, via checks every quartal.

1.1 Document Structure

The deliverable comprises the following sections:

- **Section 2 – Data Summary** provides an overview on the types of data, data collection aspects, and data formats.
- **Section 3 – FAIR Data** describes the FAIR methodology applied in CODECO for data findability, interoperability, and handling.
- **Section 4 – Allocation of Resources** covers the responsibility of Partners regarding data management and curation.
- **Section 5 – Data Preservation, Storage, and Privacy** addresses the methodology for data storage, backup, and deletion.



- **Section 6 – Ethical Aspects** covers the ethical methodology and processes followed in CODECO.
- **Section 7 – Others** summarizes any other issue at the current DME version.
- **Section 8 – Use-cases Initial Considerations** addresses initial data curation and management aspects for the CODECO use-cases.
- **Section 9 – Summary** concludes the deliverable.
- **Annex I** provides the CODECO dataset template for data collection.

1.2 Dependencies

D2 corresponds to the DME and covers data management and ethical aspects. Relevant deliverables that relate with D1 are:

- D1 – CODECO Handbook and Gender-Neutral Guidelines, M3.
- D5 – Risk Assessment and Management Plan, M3.
- D21 – Dissemination, Communication, Promotion Plan, M3.

2 Data Summary

2.1 Types of Data

Different types of data will be collected, generated, and processed in the CODECO project, including the following ones, which represent a starting point to be regularly revised in the project, and summarized in Table 1.

- **Project Management and Coordination data:** data used internally in the management of the project, including minutes, project reports, financial and progress reports, data collected on the communication and dissemination events, as well as partners contact list and project deliverables (public and confidential).
- **Experimental data:** data concerning the use-case pilots, its processing and aggregation results on the edge and communication metadata.
- **Research data:** data used to develop the CODECO technologies such as surrogate datasets used to train *Machine Learning (ML)* algorithms.
- **Derived data:** data resulting from the processing and analysis of data collected in the project, both at the Edge-Cloud as well as processed data and curated databases with the partners.



- **Source code:** CODECO will develop open-source code that will be made publicly available via the following ECL repository: <https://gitlab.eclipse.org/eclipse-research-labs/codeco-project>. This code is subject to review, updates, and overall source-code management rules. Although it is not considered directly “Data”, it is effectively the result of the work carried out in this project and needs to be placed here for consistency and discussed where deemed necessary.
- **Engagement data:** CODECO counts with a Research and Innovation Community Engagement Programme (WP7) and shall also interact with the open-source community, creating for the purpose events such as demo camps.

Table 1: Categories of datasets expected in CODECO.

Data set	Examples	WP
Project management and coordination data	Project reports, GeA meeting minutes, financial and progress reports	WP1
Experimental data	Data from use-cases, from the Innovation and Community Engagement Programme	WP3-WP5
Research data	Data used for the development and validation of CODECO technologies	WP3-WP5`
Derived data	Scientific publications, white papers, presentations, videos	WP3-WP5
Source code	CODECO toolkits and components; CODECO Apps developed for use-cases	WP3-WP6
Outputs derived from AB meetings	Meeting minutes, surveys, etc.	WP1
Outputs from the Innovation and Community Engagement Programme	Surveys, polls, minutes, proceedings	WP6
	Scientific publications, white papers, presentations, videos	WP2-WP7
Engagement data	Datasets derived from the Research and Innovation Community Engagement programme, and from additional cooperative and interaction meetings, e.g., minutes, proceedings	WP6, WP7

2.2 Data Collection

On a first level of the project, data collection includes mainly project management related data (meetings, work or visit plans, venues, and dates), architecture and frameworks documentation with versioning and traceability, meeting minutes and protocols with list of participants, and project reports / deliverables. Data will be, as far as possible, stored in widely used formats such as MS Excel (.xls/.xlsx, .csv), MS Word (.docx), image files (.jpeg, .TIFF), among others. Most files are expected to be small. CODECO does not rule out the use of datasets with richer semantics and metadata (e.g., labelled datasets) should this be required for research purposes.



Data collected at a second level of the project, corresponds to source code development. It will be collected in an on-line GIT repository at ECL Research Labs GitLab: <https://gitlab.eclipse.org/eclipse-research-labs/codeco-project>.

In the third level of the project, data is expected to be collected from different cyber physical systems (e.g., sensors) and will be organized in datasets per use-case. This corresponds to the creation of **raw data** (semi-processed data from, e.g., sensors, cameras). This data can be useful to the scientific community upon request or assessment; for industries eventually to analyse its suitability to their context. This data needs to be processed for removal of directly and indirectly identifying *personally identifiable information (PII)* before storage.

In the fourth level, derived data from the collected experimental data will be created. This refers to scientific and analytical analysis resulting from the execution of the source code (second level on top of the collected cyber-physical system data (third level). This will be relevant to scientific community as well as for use by the public, for example, when publishing news on the social media/website, or on data observatories.

2.3 Data Formats

Initial formats proposed to assist in aligning the work developed and the data curation are described in *Table 2*. The listed formats are not exhaustive.

Table 2: Initial data formats used in CODECO.

Type	Recommended formats
Project management and coordination data	Based on the project templates: .Word, docx, Excel, xlsx, ppt, open document odt
Experimental data	Sensor descriptions: WoT TD; YAML; YANG
Derived data	Latex for scientific publications; word for communication
Source code	Markdown for code documentation
Outputs derived from AB meetings	Based on the project templates: .Word, docx, Excel, xlsx, ppt, open document odt
Outputs from the Innovation and Community Engagement Programme	Based on the project templates: .Word, docx, Excel, xlsx, ppt, open document odt; xml

2.4 Template for Collecting Datasets

A template for collecting and specifying the datasets to be managed in the project is included in Appendix 1 and has also been uploaded to the CODECO repository being available to Partners under [02-WorkPackages/WP2-Use-Cases-OS Ecosystem-ArchitecturalDesign/T2.1/DataCollection](#). It prescribes collection of the following information:

- **Dataset name:** A descriptive name for the dataset.



- **Dataset unique ID:** A unique identifier for the dataset based on some naming convention. To ensure uniqueness the id should be preceded with the prefix “HEU-101092696-CODECO”.
- **Dataset description:** A short but comprehensive description of the dataset.
- **Related WP/Task:** The WP(s)/task(s) of the CODECO workplan that include the work, which will produce/generate the dataset.
- **Data origin:** Information about the activity and actor that will produce the data.
- **Will you re-use any existing data? If YES, how?** The answer to this question indicates whether the dataset is produced based on existing datasets. It also explains the mechanisms (e.g., type of processing) that enables the production of the dataset.
- **Methodologies for data collection / generation:** It illustrate the methodologies of the data collection and production.
- **Data format:** The format of the dataset (e.g., CSV, JSON, PDF).
- **Data storage:** Information about where and how the dataset will be persisted and stored. This may include storage of the dataset in some open access repository like Zenodo.
- **Expected size of the data:** An indication of the order of magnitude of the volume of data in the dataset.
- **Metadata and standards:** A list of metadata formats and/or of related standards that are used in the representation of the dataset.
- **For whom might the dataset be useful?** An indication of the stakeholder groups that will benefit from access and use of the dataset.
- **Data access, sharing and licensing:** Information about the licenses of the dataset (e.g., Creative Commons (CC)), as well as information about how to access and share the data (e.g., via some Application Programming Interface (API)).

3 FAIR Data

3.1 Making Data Findable and Openly Accessible, Including Provisions for Metadata

CODECO addresses data following the FAIR methodology in its ecosystem, e.g., across the different use-cases and on the training tools. The specific data privacy and management



aspects are therefore also related with the policies at such end-user sites. To maximize the impact of CODECO research data, the results are shared within and beyond the consortium. Selected data and results will be shared with the broad scientific community and other stakeholders, via the ECL CODECO repository, through publications in scientific journals and presentations at conferences, as well as through open access data repositories.

All the CODECO assets shall be openly available to the community: data sets, software, training, scientific publications. In what concerns software, CODECO has as one of its outcomes to provide an open-source software Edge. Therefore, the project follows good practices for the development of open-source software.

The processed data will also be publicly available via the **CODECO ZENODO community**¹. ZENODO is a general-purpose open-access repository developed under the European Open AIRE program and operated by CERN. Files will be deposited under closed, open, or embargoed access. In case some data has an embargo status, the date for the embargo will be provided and the repository will restrict access to the data until the end of the embargo period; at which time, the content will become publicly available automatically.

Project outputs, including oral presentations, posters, and scientific publications (e.g., scientific papers, book chapters), will be shared also via the ZENODO CODECO after an embargo period, if applicable. The CODECO ZENODO community is already linked to the CODECO Website and shall be shared via the CODECO social channels, so that the public and stakeholders are aware of the project outputs.

Regarding metadata, it will be recorded in a text file (e.g., a ReadMe file) to serve as support for data interpretation and deposited in ZENODO together with the data to describe.

As stated in the Grant Agreement of the CODECO project, metadata must be in a standard format and include the terms:

- “European Union (EU)” and “Horizon Europe”.
- the name of the action, acronym, and grant number.
- the publication date and length of embargo period if applicable.
- a persistent and unique identifier (such as Digital Object Identifiers (DOI)).

¹ <https://zenodo.org/communities/he-codeco/>



Note that the ZENODO platform will facilitate the project to assign and management the persistent identifiers of its datasets, notably DOIs. The latter are a key prerequisite for ensuring a minimal level of fairness for the CODECO datasets.

3.2 Making Data Interoperable

To ensure interoperability, all datasets will use standardized formats, be accessible and easily shared; and the same applies to metadata. Open formats, such as YANG, YAML, are the preferred ones. As the project progresses and data is identified and collected, further information on making data interoperable will be outlined, namely, information on data and metadata vocabularies, standards, or methodology to follow to facilitate interoperability. Moreover, CODECO aligns with the Kubernetes recommendations (open documentation, open formats, and open-source code).

3.3 Increase Data Reuse

The CODECO open ecosystem is being built on an Open Science practice, where models for co-creation, specification and validation of the research are planned. Via the establishment of open use-cases, and via the offering of open-source software stored on the ECL Research Labs GitLab, CODECO shall be open at an initial stage of specification to the broad community. ECL will store all the results on a public GitLab server, and when applicable, ECL will put available the software developed in the project in public repositories, where binaries and sources can be retrieved. All the CODECO assets shall therefore be openly available to the community: data sets, software, training, scientific publications. This shall increase impact and reusability of CODECO features.

The open datasets of the project will be licensed based on a Creative Commons (CC) license format. Business friendly licenses will be on the other hand used for the OSS results.

3.4 Initial Dataset Handling

To kick-start development, we can use initial example datasets from different sources and contexts. This data is not to be shared outside the consortium, as it not generated by CODECO research. If some of this data (augmented or modified/synthesized) does find a purpose in the continuation of the development and it becomes part of a relevant dataset, then it will be anonymised following the rules laid out in this document.



4 Allocation of Resources

4.1 Responsible Entities

CODECO counts with a hierarchical Data management structure, where INO (WP1, T1.1) represents the **Project Data Manager** (Vitor Vieira, Joana Rodrigues), assisting the Coordinator (FOR, Rute C. Sofia) in the overall data management.

All CODECO Partners are involved and are responsible, in one way or another, in data management (WP1, T1.1).

Under the guidance of INO, the CODECO partners play a major role ensuring that the data management provisions are applied.

4.2 Responsible Persons

Table 3 provides a summary concerning the data processes and responsible roles, and persons.

Table 3: Responsible Entities and Persons.

Description	Responsible	Partner and Person
Collection and curation of data in pilots	Pilot leaders	P1: UGOE, Xiaoming Fu P2: i2CAT, Rizkallah Touma P3: TID, Luis Contreras P4: UPM, David Jimenez P5: FOR, Rute C. Sofia P6: ALM, Andries Stam
Processing and preservation of data across CODECO	WPL of WP1 and WP6	Rute Sofia (FOR), Vitor Vieira (INO) and Joana Rodrigues (INO), Xiaoming Fu (UGOE)
Responsible for publishing and sharing data	Pilot leaders, data owners	P1: UGOE, Xiaoming Fu P2: i2CAT, Rizkallah Touma P3: TID, Luis Contreras P4: UPM, David Jimenez P5: FOR, Rute C. Sofia P6: ALM, Andries Stam
DMP creation and updates	Project Data Manager	Vitor Vieira (INO) and Joana Rodrigues (INO)

4.3 Curation

The main aspects of curation are as follows:

- Data from **CODECO first level** (rf. section 2.2) will be organized and preserved by the CODECOs WP1 and WP6, depending on the type of data. Project management data will be processed internally and shall not be passed on to third parties outside of the project, except funding entities as may be required by any applicable reporting



obligations. Deliverables, scientific publications, will be stored in ZENODO and linked to the Website, as has been explained in Deliverable D1.

- Data resulting from the **CODECO second level** (rf. section 2.2) will be preserved and stored via the Eclipse Research Labs GitLab: <https://gitlab.eclipse.org/eclipse-research-labs/codeco-project>.
- Data resulting from the **CODECO third level** (rf. section 2.2) will be processed and personally identifiable features (faces, license plates, etc) will be removed or obfuscated and stored in the open CODECO ZENODO community.
- Data resulting from the **CODECO fourth level** will be organized and made available both via the CODECO Website and via ZENODO.

5 Data Preservation, Storage and Privacy

The following guidelines will be followed to ensure the security of the data:

- Store project related data in the cloud at the secure EU server <https://cloud.fortiss.org/> to avoid loss of data.
- Encrypt data if it is deemed necessary by the participating researchers.
- Limit the use of USB flash drives.
- Label files in a systematically structured way to ensure the coherence of the final dataset.

5.1 Data Storage and Backup

All data detailed in section 2.1 will be stored and controlled by the CODECO Coordinator (FOR) and the INO team (CODECO management team). The CODECO cloud is regularly backed up, and Zenodo has its own backup system.

Additionally, partners are recommended to store locally their necessary work data at all four levels described in section 2.2.

5.2 Data Selection and Preservation

The data produced at the CODECO second level relates with open-source code and is therefore a primary outcome of the project, and thus very important for the broader research community. The code, kept in the ECL CODECO GitLab, will be updated on a regular basis, with a software development standard of “*commit often and early*”. This will guarantee fully documented and reproducible source code and project results.



The CODECO third level of data will be stored at ZENODO once it has been processed and ensured that it has been anonymized. This data is very important to be available for reuse in multiple domains and different applications, not yet foreseen by the project partners.

Scientific publications and corresponding datasets will be shared through ZENODO and other databases to promote the data making FAIR (see Section 3.1). The CODECO ZENODO management team (UGOE, INO, and FOR) is responsible for the data curation and preservation.

5.3 Data Destruction

The different levels of data in CODECO are expected to live after the project lifetime. However, there may be cases where data expires, or one of the Partners (or external participants in the data collection) requests deletion of the data in accordance with Article 17 of GDPR, which concerns the Right to erasure. In this case, the data responsible entity (cf. to Table 3) will handle a secure deletion process, after consultation with the CODECO Data Manager (INO) and the CODECO Coordinator (FOR). For this process, CODECO shall consider secure data erasure methods (e.g., specific software), ensuring that the deleted data will not be recoverable after destruction. Moreover, CODECO shall create an auditable report for the process.

6 Ethical Aspects

Ethical aspects in CODECO shall be handled in WP1. The data that will be collected during the CODECO project will be handled in line with the highest ethical standards and the applicable EU, international and national law on ethical principles. The CODECO data is under no specific legal requirements and will be shared after an embargo period, if applicable. No dataset containing confidential information, or any ethical or legal issues, will be deposited.

Particularly important is the CODECO third level of data, that might contain video and photography data of people. This needs to be anonymised as early as possible – preferably on the edge directly – and blur out faces and license plates. Only then can the researchers continue to process this data and make it available to external persons following FAIR principles.

Additionally, an internal *Ethics Committee (ETC)* shall be elected until M3 to support the analysis of the proposed activities regarding data management and ethics assessment



support. The procedure for reporting by the ETC is detailed in CODECO deliverable **D1 - Handbook and Gender-Neutral Guidelines** being summarized here for the sake of clarity. The ETC will continuously communicate with the CODECO Executive Board (rf. To D1), and report to the GeA findings on each plenary meeting, and on extraordinary meetings, if required. The ETC shall meet regularly with the project management team and consortium, every quartal, aligned with the plenary meetings, to oversee the right approach to the handling of the project related data. **This is further described in D1.** The ETC will follow the proposed DME and ensure that partners follow the proposed rules, making sure that they conform to the legislation regulations in force in the countries where the research will be carried out, as well as to the EC Ethical Legislation.

6.1 Processes

The verification process is of the responsibility of the entity owning the data, as has been detailed in Table 3. The verification process is the following:

1. Identify the dataset that is relevant. (Responsibility: developer of the organization working with the dataset).
2. Verify if there is any PII present - for CODECO we adhere to directive 95/46/EC¹ which includes any information on the dataset or metadata that includes information capable of identifying a person via an ID number, or factors specific to physical, physiological, mental, economic, cultural, or social identity. (Responsibility: developer of the organization working with the dataset).
3. Removal of PII (Responsibility: developer of the organization working with the dataset).
4. Request an additional verification of the data by a responsible from a different organization in the consortium. (Responsibility: developer of the organization working with the dataset contacts a second opinion to verify if all PII was removed from the dataset).
5. The dataset is made available to the ETC for verification of compliance, and any other potential ethical issues and bias. (Responsibility: ETC).
6. The controller of the data for publishing (see section 2.2), verifies the compliance of the dataset as well with the set rules and then publishes it.

All these numbered steps need to be documented and confirmed by the persons involved in these steps, and this document is attached to the dataset prior of publishing.



6.2 GDPR Compliance

The consortium is committed to conducting the project in accordance with any applicable local laws that apply such as the General Data Protection Regulation 2016/679 (GDPR) (collectively referred to as the EU Data Protection Legislation).

To ensure GDPR compliance, no data will be collected or used without the explicit informed consent of the subjects (e.g., users) concerned. In this direction, users/subjects provided data must sign a proper informed consent form. The form will make explicit who has access to the collected data, as well as for how long the data will be securely stored and subsequently deleted. Users/subjects will be also free to access their own data whenever they like. This will empower the Individuals involved to make a voluntary informed decision about whether to participate in CODECO processes based on knowledge of the scope, purpose, procedures, and outcomes of the project. Moreover, CODECO actors (e.g., end-users) will also have the possibility to gain access to additional information about their personal data and the way they are processed. Likewise, they will be able to revoke their consent at any time.

7 Other Issues

At this moment, no other national / funder / sectorial / departmental procedures for data management have been applied.

8 Use-cases Initial Considerations

The CODECO key technological blocks will be explored via the use of multiple use-cases (UC) which will serve as basis for experimentation and demonstrations, focus on the development of 6 UC across 3 different European competitiveness markets: Smart Cities, Energy, Manufacturing. Each partner leading a UC will be its Data Manager. A specific team member of the partner will oversee data gathering and quality management, aligned with WP1. Each UC leader will therefore execute its own data management on the specific UC. In this section, initial considerations concerning data aspects are done for each use-case. **These initial considerations will have to be revised for the next version of the DME, once the use-case leaders define the datasets.**

8.1 Smart Cities

P1– Smart Monitoring of the Public Infrastructure – led by UGOE, will be employed to assist the smart monitoring of Göttingen’s public infrastructure by integrating CODECO in mobile Edge nodes across the city. Similarly, **P2 - Vehicular Digital Twin for safe urban mobility** – led by I2CAT, demonstrate how the CODECO framework can support the



deployment of vehicular digital twin services that enhance safety of Vulnerable Road Users (VRU) in urban environments. Both UCs have the potential to film people and cars, which might constitute a privacy violation when storing the collected and processed data. For this purpose, CODECO partners responsible for the use case will immediately blur faces and license plates of people and vehicles filmed at earliest Edge iteration as possible. Stored data will only contain this censored / anonymized data version.

Concerning **P3 – MDS across Decentralised Edge-Cloud** – led by TID, focus on the smart and efficient distribution of media content across a multi-domain, the CODECO project will also need to consider privacy concerning to data collection and strict data privacy policies and regulations, such as the GDPR in the European Union. This may include measures such as anonymizing user data and other Personally Identifiable Information present (PII), obtaining user consent for data collection, and implementing robust data security measures to protect against unauthorized access and data breaches.

8.2 Energy

Like Smart Cities UCs, **P4 – Demand Side Management in Decentralised Grids** – led by UPM, focuses on the deployment of Edge-based Smart Energy Nodes to provide a solution for energy controlling and adjusting in real time, people that are filmed / photographed will be immediately blurred with the same process. Stored data will only contain this censored / anonymized data version. The data is not expected to contain PII, but if it does contain it, it will also be removed.

The use-case will handle energy datasets and if consumable energy datasets will be used, a consent form provided by CODECO shall be collected.

P6 – Smart Buildings – led by ALM, will apply the CODECO technologies to their Crownstone smart buildings solution. Like happens in the other UCs, people that are filmed / photographed will be immediately blurred with the same process. Stored data will only contain this censored / anonymized data version.

8.3 Manufacturing

P5 – Wireless AGV control for flexible factories – led by FOR, focuses on the optimization of the overall efficiency of AGV. If people are involved, people that are filmed / photographed will be immediately blurred with the same process. Stored data will only contain this censored/anonymized data version. data must be properly managed and stored to prevent data breaches or loss of sensitive information. Additional data that may be collected may relate with maps of obstacles for each AGV, or additional tasks passed to the AGVs.



9 Summary

This document describes the main principles and guidelines for data management under the CODECO project. As a living document, it will be updated throughout the project's lifetime. Further versions of the DME will include the eventually updating of the online research data repository where data are collected and shared, and the description of the dataset and research data gradually generated and collected. The next update of the DME will include a concrete list of datasets that will be produced by project, including detailed information about each one of the datasets.



Annex 1 – Dataset Information Collection Template

Dataset Information Template (completed by each partner in charge of the dataset)

Dataset name	
Dataset unique ID	
Dataset description	
Related WP/Task	
Data origin	
Will you re-use any existing data? If YES, how?	
Methodologies for data collection / generation	
Data format	
Data storage	



Expected size of the data	
Metadata and standards	<i>(e.g., Basic descriptive metadata (such as title, author, date created and keywords) will accompany the dataset)</i>
For whom might the dataset be useful?	
Data access, sharing and licensing	

ⁱ Directive 95/46/EC <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:l14012>

